January 2015

Creation and Validation of the Critical Thinking about Sustainability Scale

Vivian Gregory Alexander
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Vivian Alexander

Entitled
THE CREATION AND VALIDATION OF THE CRITICAL THINKING ABOUT SUSTAINABILITY SCALE

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Yukiko Maeda
Chair

Aman Yadav

Loring Nies

Chantal Levesque-Bristol

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Yukiko Maeda

Approved by: F. Richard Olenchak

Head of the Departmental Graduate Program Date

DEVELOPMENT AND VALIDATION OF THE CRITICAL THINKING ABOUT SUSTAINABILITY SCALE

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Vivian G Alexander

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2015

Purdue University

West Lafayette, Indiana

I would like to dedicate this work to my mother, Althea Isidore, who was both mother and father to my sister and I. Without, her great sacrifice, this moment would not be possible. I would also like to dedicate this work to the love my life, Tennille Auguste, who worries about me more than I worry about myself.

ACKNOWLEDGEMENTS

This work and being able to complete this work would not be possible without the guidance and support of many persons. First, I would like to express my sincere thanks to my undergraduate mentor, Dr. Robert Vanderburg for engaging me in educational research from the very first week that I entered my undergraduate education program. His continued mentorship and guidance were critical to my development as a researcher capable of completing a dissertation. Next, I would like to express gratitude to my academic advisor, Dr. Yukiko Maeda who showed great patience and understanding with me throughout my time at Purdue. She was always available to provide mentorship and motivation especially as I was adapting to a completely new cultural experience.

I also extend my gratitude to my advisory committee who were supportive of my study and made every effort to make this study possible. I would not have been able to make it this far without my friends at Purdue who quickly became like family. Our constant support for each other during the good times and the bad, helped make life as a graduate student a little more bearable. Last but not least my family who were as deeply invested in this process as I was.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

ABSTRACT ix

CHAPTER 1. INTRODUCTION 1

1.1 Research Problem 4

1.2 Purpose of Study 7

1.3 Research Questions 8

1.4 Significance of Study 8

CHAPTER 2. LITERATURE REVIEW 10

2.1 Sustainability 10

2.1.1 Ecological Sustainability 11

2.1.2 Economic Sustainability 12

2.1.3 Social Sustainability 15

2.1.4 Sustainability and Sustainability Development 18

2.2 Critical Thinking 19

2.2.1 Philosophical Approach to Critical Thinking 19

2.2.2 Psychological Approach to Critical Thinking 20

2.2.3 Educational Approach to Critical Thinking 20

2.2.4 Critical Thinking and Domain Specificity 21

2.2.5 Critical Thinking Transfer between Domains 23

2.2.6 Critical Thinking Disposition 24

2.2.7 Operationalizing Critical Thinking 25

CHAPTER 3. METHODS 29

3.1 Exploration of Experts' Conceptualization of Sustainability 29

3.2 Scale Format Selection 32

Page
3.3 Item Development and Scoring 34
3.4 Evaluation of the Content Validity of the CTSS 36
3.5 Evaluation of Construct Validity..36
3.5.1 Sample and Sampling Procedure 38
3.5.2 Scale Administration 39
3.5.3 Evaluation of CTSS Dimensionality 40
3.5.4 Evaluation of CTSS Question Quality..43
CHAPTER 4. RESULTS 47
4.1 Experts' Conceptualization of Sustainability.......................................47
4.1.1 Variations in experts' conceptualization of sustainability.......................48
4.1.2 Sustainability vs. sustainable development 52
4.1.3 Students' knowledge of sustainability 54
4.1.4 Measuring students' knowledge of sustainability....................................55
4.2 Evaluation of Construct Validity..60
4.2.1 Scoring Reliability and Score Distribution...60
4.3 Evaluation of CTSS Dimensionality with Exploratory Factor Analysis.....66
4.4 Evaluation of CTSS item quality with item response theory 77
4.4.1 Thresholds and slopes..78
4.4.2 Item information function (IIF) evaluation...84
4.4.3 Test information function (TIF) evaluation 87
CHAPTER 5. DISCUSSION.. 92
5.1 Experts' Conceptualization of Sustainability.......................................93
5.2 Evaluation of CTSS Construct Validity and Quality 97
5.3 Limitations and Future Research...107
5.4 Conclusion...109
LIST OF REFERENCES 111
APPENDICES
Appendix A Sustainability Scales 128
Appendix B Critical Thinking Scales 129

Page
Appendix C Interview Protocol 130
Appendix D Permissions 131
Appendix E Scenario 1 133
Appendix F Scenario 2 134
Appendix G Scenario 3 135
Appendix H Critical Thinking about Sustainability Rubric 136
Appendix I Category Characteristic Curves for KFCC Scenario 138
Appendix J Category Characteristic Curves for BBD Scenario 141
Appendix K Category Characteristic Curve for OTO...................................144
VITA 147
PUBLICATIONS... 152

LIST OF TABLES

Table Page

Table 3.1 39

Table 4.1 57

Table 4.2 61

Table 4.3 63

Table 4.4 65

Table 4.5 66

Table 4.6 68

Table 4.7 69

Table 4.8 69

Table 4.9 72

Table 4.10 73

Table 4.11 78

Table 4.12 80

Table 4.13 91

Table 5.1 105

Table 5.2 106

Table 5.3 106

Figure Page
Figure 4.1 70
Figure 4.2 70
Figure 4.3 71
Figure 4.4 85
Figure 4.5 86
Figure 4.6 87
Figure 4.7 89
Figure 4.8 89
Figure 4.9 90

ABSTRACT

Alexander, Vivian G. Ph.D., Purdue University, December 2015. Creation and Validation of the Critical Thinking about Sustainability Scale. Major Professor: Yukiko Maeda.

Critical thinking and Sustainability are both key topics of interest to higher education institutions. Over the decades, researchers have sought to measure undergraduate students' critical thinking ability as well as their understanding of sustainability. However, while researchers acknowledge the importance of critical thinking to approaching the subject of sustainability in undergraduate students, they have given little attention to measuring undergraduate students' ability to think critically about sustainability. Therefore, this study utilizes a sequential mixed-methods design to create and validate a scale to measure undergraduate engineering students' ability to think critically about sustainability called the Critical Thinking about Sustainability Scale (CTSS).

CHAPTER 1. INTRODUCTION

Balancing the demands for raw materials, food, clothing, shelter, energy, and other goods, with the environmental limits of ecosystems is a constant concern throughout human history (Ponting, 2007). Researchers consider sustainability as the best way to address the vast, complex, and interrelated environmental and societal problems that affect both current and future generations (Waas, 2011). A concern for sustainability arose in the early seventies as a growing number of people realized that the degradation of the environment seriously undermined our ability to ensure expanding prosperity and economic justice (Clugston & Calder, 1999). Sustainability describes the delicate balance between the economic, environmental, and social health of a community, nation, and the earth (Fricker, 1998; World Commission on Environment and Development [WCED], 1987). The preservation of this delicate balance involves meeting human needs by providing an ecologically stable and healthy environment and by addressing the social and cultural needs of people (Littig & Grießler, 2005).

Economic sustainability describes the “maintenance of capital” or “non-declining capital” (Goodland & Daly, 1996). Economic sustainability focuses on the renewable (e.g., forests) and exhaustible (e.g., minerals) physical inputs into the production process (Goodland, 1995). Economists view sustainability not as the standard that guides global development but as one element of a desirable development path. Ecological sustainability focuses on sustaining global life-support systems indefinitely (Goodland, 1995). Ecological sustainability seeks to preserve the sources of raw materials used for human needs and ensuring that the sinks for human wastes are not exceeded (Goodland & Daly, 1996). Social sustainability is a quality of societies that signifies the nature-society relationship mediated by work and the relationships within the society (Littig & Grießler, 2005).

According to Littig and Grießler (2005) social sustainability occurs when a society and the institutional arrangements within a society: (1) satisfy an extended set of human needs, (2) preserves the reproductive capabilities of nature over a long period of time, and (3) the normative claims of social justice, human dignity and participation is fulfilled. Although sustainability is most prominently associated with the ecological crisis phenomena, studies of ecological sustainability lie in a society-oriented definition of problems. Ultimately, environmental sustainability relates to problems of social justice, gender-equality, and political participation (Becker, Jahn, & Stieß, 1999). This makes sustainability quite a challenge for the social sciences, as sustainability research is not just about ‘natural’ processes but also about understanding the social processes that affect society’s interactions with nature (Littig & Grießler, 2005).

There are many challenges that universities face on the path to sustainability. One challenge is how much change is required in university curricula to accommodate sustainability. The lack of a clear understanding of sustainability and sustainable development makes the integration of sustainability principles in university curricula difficult (Thomas, 2009). Another challenge affecting the integration of sustainability principles into university curricula is that sustainability involves merging the largely different value sets of economics, environment, and the society (Chau, 2007). Moreover, researchers and stakeholders on the global, local, and community level each have their own conceptualizations of a sustainable society (Thomas, 2009). This makes the integration of sustainability principles into university curricula even more challenging. Therefore, researchers need a different process to engage with the many issues associated with sustainability and teaching sustainability principles (Thomas, 2009).

Critical thinking (CT) has a central role in achieving sustainability and teaching sustainability principles (Saravanamuthu, Brooke, & Gaffikin, 2013). According to Matsuura (2007), sustainability presents a challenge of learning how to live differently. It involves asking how we are to raise future generations with values, attitudes, and understandings different from our own. Education is key to this, but it is much more than including education for sustainable development into the curriculum and teaching materials. Education about sustainability is also about cultivating capacities of critical thinking, careful analysis, respect for others and forward-thinking abilities, which enable people to reflect upon and change their behavior, values, and life-styles (Ferrer-Balas, Lozano, Huisingh, Buckland, Ysern, & Zilahy, 2010). Critical thinking is important for sustainability and sustainability education for several reasons: it encompasses the ability to recognize an existing problem, seeking proof for evidence provided, verifies the accuracy of this proof, and makes use of this knowledge when making decisions (Ozturk, Muslu, & Dicle, 2008).

Enhancing student’s CT skills is an important goal of modern education, as it equips students with the competency necessary to reason about social affairs in a rapidly changing world (Halpern, 2001). Developing such competency requires that students go beyond absorbing textbook knowledge to developing skills in judging information, evaluating alternative evidence and arguing with solid reasons. Students need these CT skills to perform well in school, in their future workplaces, and in social and interpersonal contexts where they need to make sound decisions carefully and independently on a daily basis (Ku, 2008). In many countries around the world, CT is a major educational goal (Halpern, 2001). For example, education reports from countries such as the United States, United Kingdom and Australia, list CT as a key area for development and assessment in higher education (Miller & Leskes, 2005).

1.1 Research Problem

Critical thinking is a widely used term that includes skills in applying, analyzing, synthesizing, and evaluating information and the disposition to apply these skills (Facione, 1990). Earlier definitions of critical thinking emphasized the cognitive component of CT—that it is a skill, mental procedure, or rationality (Ennis, 1962; McPeck, 1981). Later conceptualizations of CT included an intentional and motivational aspect described by scholars as a critical thinking disposition (Facione, 1990; Halpern, 1998). Since the construct of CT is abstract and multi-natured, its assessment has also been indefinite (Paul, 1993). Currently, there is little consensus among researchers on how critical thinking should be measured and over the decades, researchers have developed a number of CT instruments. These include, but are not limited to the Watson–Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) and the California Critical Thinking Skills Test (Insight Assessment, 2013). However, previous studies utilizing these different instruments as estimates of individuals’ critical thinking competence assume that the chosen measurements of critical thinking are compatible with the conceptualization of critical thinking.

While these CT tests overlap in some aspects, they vary in their purpose, format, and context (Ku, 2009). Although they assess CT in a more diverse range of domains, both scales have two common characteristics: 1) they do not focus on measuring students' ability to think critically about sustainability and 2) it may be financially impractical for researchers to use these scales for a large number of students. For example, the Watson-Glaser critical thinking assessment (1980) Form A test consisting of 25 booklets costs approximately 724 U.S. dollars and this may prove financially impractical for some researchers to administer to a large number of students.

Currently, HEIs are integrating sustainability principles in all aspects of university life, including teaching, learning, research, and societal integration (Du, Su, & Liu, 2013). Incorporating sustainability into higher education is essential for students to understand how their everyday actions can affect the relationship among environmental, economic, and social issues and to influence their actions as local and global citizens (Moore, 2005). As mentioned before, CT is one of the important aspects of sustainability education, as it encompasses the ability to recognize an existing problem, to seek proof of the evidential and to gather knowledge about the accuracy of this proof and make use of this knowledge (Ozturk et al., 2008). Critical thinking has become an integral part of higher education (Halpern, 2001) and understanding sustainability requires a focus on competencies and higher order thinking skills (Wagner & Dobrowolski, 2000). The challenges associated with achieving sustainability are complex and multifaceted and require problem-solving methods to resolve the opposing forces of seemingly intractable issues (Holden et al., 2008). Moreover, a number of environmental questions lie at the root of critical thinking (Tilbury, 1995).

Researchers recognized that a transition to sustainability from societies' current ways of operating requires institutional, social, and individual change due to the inherent characteristics of sustainability: it is fundamentally an applied, problem-based concept rather than a purely theoretical one; it integrates many domains of knowledge and practice, and utilizes collaborative approaches (Robinson, 2004). Therefore, higher education institutions have a responsibility to develop in their students the ability to analyze information, construct arguments, and act with a high degree of autonomy and self-determination (Wals & Jickling, 2002). Higher education institutions must also develop in their students the ability to cope with uncertainty, poorly defined situations and conflicting norms, values, interests and reality constructions (Wals & Jickling, 2002). These institutions need to train students to become responsible leaders who consider the social, economic, and environmental factors in making decisions (Locke et al., 2009) and to develop students' CT abilities (Toutkoushian, 2005).

In order to incorporate sustainability principles into higher education researchers and institutions have developed tools for measuring students' knowledge about sustainability. For example, Kagawa (2007) developed a scale to explore undergraduate students at the University of Plymouth current understandings and perceptions of and attitudes towards sustainable development and related concepts. Similarly, Azapagic, Perdan and, Shallcorss (2005) developed an international survey to investigate engineering students' knowledge of sustainable development.

Yet, despite the efforts to measure both sustainability and CT, there are several challenges that researchers face. First, much of the developed sustainability scales focus mainly on the ecological component of sustainability. Second, many of the sustainability scales measure students' knowledge or perceptions about sustainability and not students' ability to think critically about sustainability. Third, the high cost of the current CT assessments coupled with their lack of emphasis on sustainability principles precludes the use of these assessments as measures of undergraduate students' ability to think critically about sustainability. Therefore, there is a need for a tool to measure university students' ability to think critically about the three aspects of sustainability.

1.2 Purpose of Study

Therefore, the goal of this study is to develop a reliable and valid scale to measure undergraduate engineering students' ability to think critically about sustainability, called the Critical Thinking about Sustainability Scale (CTSS). I focused on engineering students for this study because a rapidly rising global population and improving standards of living are challenging engineers to use the limited natural resources of the world to satisfy ever-increasing human demands (Davidson et al., 2010). However, despite the increasing demands on engineers, few engineering schools have made major updates to their courses and curricula over the past few decades to help their graduates deal with the increasing demands of the profession (Davidson et al., 2010).

This study will provide researchers with a scale that measures undergraduate students' ability to think critically about sustainability, thereby filling the identified gaps in the past literature. To achieve this goal, I utilized a sequential mixed methods design consisting of a qualitative phase followed by a quantitative phase. The qualitative phase consisted of one-to-one interviews with experts on the topic of sustainability and served two purposes: 1) to gain knowledge about their conceptualization of sustainability, and 2) to integrate their knowledge with the current literature to develop an operational definition of sustainability. The quantitative phase consists of the development and psychometric evaluation of the CTSS. I selected undergraduate students because the vast majority of research on sustainability and engineering occurs at the undergraduate level.

1.3 Research Questions

Two major research questions guide the development of this study. The first research question is qualitative in nature and the second research question is quantitative in nature.

1. How do sustainability experts conceptualize sustainability?
    a. To what extent is sustainability experts' conceptualization of sustainability consistent or different among each other?
2. To what extent does the CTSS measure students' ability to think critically about the ecological, economic, and social aspects of sustainability?
    a) To what extent is the proposed scale reliable and valid?

1.4 Significance of Study

Despite the importance of higher education institutions to creating a sustainable world and promoting students' CT abilities, there seems to be no literature that addresses the need to measure students' ability to think critically about sustainability issues partly due to the limited availability of such measurements. Additionally, available sustainability measures defined the concept narrowly and tended to have weak evidence of psychometric properties. Therefore, the development of a new scale focusing CT in a sustainability context with strong psychometric evidence will provide a significant contribution to research and practice by providing a medium to measure undergraduate students' ability to think critically about sustainability. Additionally, the qualitative aspect of the study will generate new knowledge of defining sustainability by integrating sustainability experts' conceptualizations of sustainability with existent sustainability literature.

CHAPTER 2. LITERATURE REVIEW

The literature review contains two sections. The first section of the chapter involves a review of the literature on sustainability in order to provide a conceptual framework of the topic. To this end, the review of literature on sustainability describes the ecological, economic, and social aspects of sustainability. The second section provides a review of literature concerning critical thinking, also with the aim of developing a conceptual framework. The literature review on critical thinking covers the different aspects of critical thinking including the disposition to think critically.

2.1 Sustainability

There is little dispute that our present path is unsustainable (Fricker, 1998). In the early seventies people realized that the degradation of the environment would seriously undermine our ability to ensure expanding prosperity and economic justice (Clugston & Calder, 1999). Thus, sustainability arose from the recognition that the profligate and inequitable nature of current patterns of development, will lead to biophysical impossibilities in the future (Goodland & Daly, 1996). The publication of the Brundtland Report (WCED, 1987) and Talliores Document (1990) are global expressions of concern about the environment. However, the challenge of sustainability is neither wholly technical nor rational, but instead, requires a change in attitude and behavior (Fricker, 1998).

Researchers describe sustainability as the delicate balance between the economic, environmental, and social health of a community, nation, and the earth (Fricker, 1998; WCED, 1987). Preserving this delicate balance hinges on the idea of meeting human needs by the providing an ecologically stable and healthy environment, and by addressing the social and cultural needs of the people (Littig & Grießler, 2005). While there is little dispute that sustainability comprises: ecological, economic, and social aspects, researchers differ in their views on the definition of sustainability (Littig & Grießler, 2005). Of the three aspects of sustainability, the operationalization of ecological and economic sustainability are the least disputed, while there remains much more disagreement concerning the operationalization of social sustainability (Omann & Spangenberg, 2002). The subsequent sections will provide an overview of the three aspects of sustainability beginning with the ecological aspect of sustainability, followed by economic sustainability, and finally the social aspect of sustainability.

2.1.1 Ecological Sustainability

The environment has now become a major constraint on human progress (Goodland, 1995). Although ecological sustainability is necessary for human life and emerged due to social concerns, the goal of ecological sustainability is to improve human welfare and society by protecting the sources of raw materials used for human needs and ensuring that the sinks for human wastes are not exceeded (Goodland & Daly, 1996). Thus, the goal of ecological sustainability is to sustain global life-support systems indefinitely (Goodland, 1995). In order to sustain global life-support systems indefinitely humanity must learn to live within the limitations of the biological and physical capacities of the environment, both as a provider of inputs “sources” and as a “sink” for wastes (Serageldin, 1993). This means that the assimilation of waste occurs within the capacity of the environment without impairing it, and the harvest rates of renewables are within regeneration rates (Goodland & Daly, 1996). In the case of non-renewable resources, holding depletion rates equal to the production rates of renewable substitutes will create a quasi-ecological sustainability (El Serafy, 1991).

Ecological and economic researchers have differing views of ecological sustainability. Ecologists have expressed concern that the world is hurtling away from environmental sustainability (Hardin, 1993; Meadows, Meadows, & Randers, 1992). In contrast, economists who have yet to reach consensus on the issue, do not generally hold this view (Goodland & Daly, 1996). However, what is not contestable is that the current modes of production prevailing in most parts of the global economy are causing the exhaustion and dispersion of a one-time inheritance of natural capital, such as topsoil, groundwater, tropical forests, fisheries, and biodiversity (Goodland & Daly, 1996). The rapid depletion of these essential resources, coupled with the degradation of land and atmospheric quality, shows that the human economy as currently configured is already inflicting serious damage on global support ecosystems, and future potential biophysical carrying capacities are probably being reduced (Daily & Ehrlich, 1992; Daily, Ehrlich, & Ehrlich, 1996).

2.1.2 Economic Sustainability

Economic sustainability is described as the “maintenance of capital” or as “non-declining capital” (Goodland & Daly, 1996). Economic sustainability focuses on the renewable (e.g. forests) and exhaustible (e.g. minerals) physical inputs into the production process (Goodland, 1995). Economists view sustainability not as that the standard that guides global development, but as one element of a desirable development path. To the economist, sustainability is maintaining dynamic efficiency and intergenerational equity (Stavins, Wagner, & Wagner, 2003). Economic sustainability has three main characteristics or assumptions: substitutability, intergenerational equity, and economic efficiency. First, the principle of substitutability implies that manmade resources are perfect and unlimited substitutes for natural resources (Hartwick, 1977; Munda, 1997; Pezzey & Toman, 2002). Second, economic sustainability shows a willingness to ignore intra-generational equity i.e., equity between members of the same generation (Toman, 2006). Third, economic sustainability espouses the principle of economic efficiency i.e., maximizing profits and minimizing losses. Researchers strongly criticize the economic approach to sustainability for placing dollar values on environmental benefits, and for the belief that manmade products are perfect and unlimited substitutes for natural resources and for the ethical issues concerning using efficiency as the sole determinant of the allocation of environmental assets (Soderbaum, 1987).

One, researchers criticize the economic approach to sustainability for placing dollar values on environmental benefits. Foy (1990) asserts that it is not feasible to provide meaningful monetary measures for all environmental benefits, because humanity is ignorant of the instrumental value of environmental functions and other life forms. Costanza and Daly (1987) argue that attempts to monetize environmental resources are subject to "quantitative bias" which gives less priority to difficult-to-quantify resources and more priority to easily quantified resources. This is problematic because the most important environmental benefits may be the most difficult to monetize. Foy (1990) asserts that it is virtually impossible to monetize the contribution of each gradual environmental change to a potential economic calamity, and therefore the monetary benefits and costs should not determine the allocation of environmental assets.

Two, researchers criticize the economic approach to sustainability for its principle of substitutability which is strongly related to the first criticism. The principle of substitutability posits that manmade resources are perfect and unlimited substitutes for natural resources (Munda, 1997; Pezzey & Toman, 2002). Therefore, the degradation of natural resources is not a concern, so long as there are offsetting increases in other forms of capital such that overall well-being can be maintained or increased over time (Pezzey & Toman, 2002). However, manmade capital is reliant on natural capital; since resources are required to manufacture capital goods, the amount of input resources required for an increase in production of substitute capital limits the production rate of the substitute resource (Munda, 1997). In other words, the amount of natural resources required to make a product limits the creation of the product. Additionally, natural resources perform important roles that go beyond the raw input for manmade goods therefore there is a limit to the substitutability between manmade capital and natural resources, as manmade goods do not possess the same multi-functionality of natural resources (Pezzey & Toman, 2002). The possibility of no substitutes for some types of natural resources also weakens the case for using monetary measures of environmental benefits to determine if environmental assets should be preserved (Foy, 1990).

Three, researchers criticize the economic approach to sustainability for the assumption that efficiency should be the sole determinant for the allocation of environmental assets among human generations and other species. The efficiency criterion assumes that the "right" decision is the one that maximizes monetary benefits. However, Kelman (1981) argued that this particular criterion of right is very controversial in moral philosophy and in public policy. The notion of human rights advocates for the treatment of people in a certain way even if the monetary benefits do not outweigh the costs and this notion directly contradicts the efficiency criterion; thus, the efficiency criterion should not be the only method used to make environmental decisions that may significantly affect future generations and other life forms (Foy, 1990).

2.1.3 Social Sustainability

Social sustainability describes the qualities of societies that signify the nature-society relationship mediated by work and the relationships within the society (Littig & Grießler, 2005). Social sustainability emerges when the work within a society and related institutional arrangements, satisfy an extended set of human needs preserves the reproductive capabilities of nature over a long period of time; and fulfils the normative claims of social justice, human dignity, and participation (Littig & Grießler, 2005). The work within a society includes the provision of education, skills, experience, consumption, income, and employment, while institutional arrangements include democracy and participation, distributional and gender equity, and independent and pluralistic sources of information (Omann & Spangenberg, 2002). Defining and assessing social sustainability faces several challenges. These include a lack of conceptual clarity about social sustainability, the inability of current institutional settings to manage the complexity of social sustainability, the lack of discussion about the interaction between the social, environmental, and economic objectives, and the tendency of researchers to suggest indicators for small sections of social development but make none for normative trends (Omann & Spangenberg, 2002).

Unfortunately, despite the considerable academic and political attention paid to sustainability and sustainable development, there is a feeling that the social aspect of sustainability is not seen to be equal to, or as important as either the economic or environmental aspects (Cuthill, 2010). International organizations lament the poor state of the social aspects of sustainability. For example, the United Nations (1993) describes social sustainability as an independent dimension of sustainable development that is equally important as the economic or environmental dimension but still lacks broad recognition by scientists as well as by decision makers. The Organization for Economic Cooperation and Development (OECD) argued that countries/institutions treat social sustainability as the social implications of environmental politics, but not as an equally constitutive component of (2001). As a result, social sustainability is the least conceptually developed of the three aspects of sustainability (Littig & Grießler, 2005).

As a means of assessing the state of social sustainability, some researchers have developed goals or indicators. For example, the Work and Environment project suggests the following criteria for defining social sustainability: self-determined life-style, satisfaction of basic needs, a reliable and sufficient social security system, equal opportunities to participate in a democratic society, and enabling social innovation (Hans-Boeckler-Foundation, 2001). Becker, Jahn, and Stieß (1999) described sustainability as a concept that consists of an analytical, normative, and political aspect and any discussion of the concept of social sustainability has to acknowledge the close and complex links among these three concepts. Analytically, the concept of sustainability implies that societal development is inseparable from its natural prerequisites. Therefore, sustainability does not support the equivalence between development and economic growth, questions the assumption of a continuous, linear, and harmonious development path for societies, and emphasizes a diversity of paths that societies can take for societal transformation (Becker et al., 1999). The analytic approach should help examine the social structures and processes that influence the exchange between society and nature (Fischer-Kowalski & Haberl, 1993).

Normatively, sustainability strives for intergenerational equity. This implies a hierarchical interdependence between the economy, society and the natural environment. Although it is possible for societies to survive without an economy, neither society nor the economy can survive without a natural environment. Therefore, social and ecological constraints take priority over economic processes. In addition to intergenerational equity, intra-generational equity as it relates to social justice, equity in gender relations and democratic participation in decision-making processes is essential for the equitable access, distribution, and management of natural resources and services (Becker et al., 1999). Politically, achieving sustainability involves reshaping the way people interact with their environment and this requires strong commitment to action. The political aspect of sustainability refers to the process of renegotiating the goals of future societal development and establishing a system of governance capable of appropriately developing and implementing policies that can achieve sustainability on a local, national, regional, and international level. This process of re-negotiation would involve three steps. Step one, the identification, and implementation of policy goals that encompass the compatibility of economic and environmental targets, equity, and social justice, and encourages the broad participation of stakeholders. Step two, a critical re-evaluation and assessment of existing institutions and institutional arrangements, and the possible actors and conflicts among them. Step three, the consequences associated with the implementation of sustainability-related strategies.

Therefore, although sustainability is most prominently associated with the ecological crisis phenomena, studies of ecological sustainability lie in a society-oriented definition of problems. Consequently, sustainability addresses the question of how societies can change to ensure intergenerational equity i.e., equity between current and future generations. Ultimately, environmental sustainability relates to 'internal' problems of social structure, such as social justice, gender equality, and political participation (Becker et al., 1999). This makes sustainability quite a challenge for the social sciences, as sustainability research is not just about 'natural' processes but also about understanding social processes that concern society's interactions with nature (Littig & Grießler, 2005). Appendix A provides a summary of Sustainability scales found through a literature search.

2.1.4 Sustainability and Sustainability Development

Sustainability and sustainable development are two related, but distinct concepts (Diesendorf, 2000). The Brundtland Report defined sustainable development as "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development [WCED], 1987). This definition describes sustainability in the context of reaching long-term goals, and achieving equity between the present and future generations; thus, sustainability is the outcome of the sustainable development process

(Diesendorf, 2000). Sustainable development (SD) should integrate social, environmental, and economic sustainability and use these three to make development sustainable (Goodland, 1995).

2.2 Critical Thinking

Over the past three decades, researchers have come to a growing agreement that education is a process of inquiry, learning, and thinking rather than an accumulation of disjointed skills and senescent information (Facione, 1990). The ability to think critically is the most desirable outcome of undergraduate education (Halpern, 2001). As a result, increasing attention has been given to the concept of CT. Critical thinking has its roots in fields of philosophy, psychology and education (Sternberg, 1986). However, despite the widespread recognition of critical thinking as an important goal of education, there is a notable lack of consensus regarding the definition of critical thinking (Lai, 2011). Appendix B provides a list of CT scales found through a literature review.

2.2.1 Philosophical Approach to Critical Thinking

The philosophical approach to CT focuses on thought processes of the ideal critical thinker. The philosophical approach describes the thinking processes of the person rather than the overt behaviors demonstrated by the critical thinker (Lewis & Smith, 1993). The philosophical approach to CT also emphasizes qualities or standards of thought (Lai, 2011, p.5). Further, the philosophical approach to CT generally focuses on the application of the rules of logic to determine the most suitable response or outcome (Lewis & Smith, 1993). However, researchers criticize this approach to CT for being unrealistic and not representative of how people actually think (Lai, 2011; Sternberg, 1986).

2.2.2 Psychological Approach to Critical Thinking

The psychological approach to CT differs to the philosophical approach to CT in two ways. First, the psychological approach describes CT within the context of the person and the environment (Sternberg, 1986, p.5). This stands in contrast to the philosophical approach that focuses on the ideal critical thinker. Second, in the psychological approach, researchers define CT by the types of actions or behaviors critical thinkers can do (Lai, 2011, p.7). However, philosophers criticize the psychological approach for oversimplifying the evaluation of CT and a lack of transfer between research and practice (Sternberg, 1986, p.6).

In one example, Bailin (2002) argued that it is a fundamental misconception to view critical thinking as a series of discrete steps or skills, and that this misconception stems from researchers' need to define constructs in directly observable ways. The argument is based on the fact that thought processes are unobservable, and psychologists focus overt behaviors which are the products of a person's thought processes (e.g., analysis, interpretation, formulating good questions). Moreover, philosophers caution against confounding CT with the skills that comprise it because CT is more than simply a collections of skills that people utilize (Facione, 1990; Van Gelder, 2005). Philosophical proponents of CT also argue can engage in the procedural aspects of CT without actually utilizing CT (Bailin, 2002).

2.2.3 Educational Approach to Critical Thinking

The educational approach to critical thinking is based primarily on classroom observation, analysis of texts, and the analysis of the thinking process (Sternberg, 1986, p.6). Bloom's (1956) taxonomy of cognitive skills is one of the most widely cited sources for educational practitioners when it comes to teaching and assessing higher-order thinking skills (Lai, 2011). The strength of the educational approach to CT is that it is derived from classroom observations and experience (Lai, 2011). However, educational theories of CT are weakened a lack of epistemological clarity characteristic of the philosophical and psychological approaches to CT, making the educational theories of CT difficult to evaluate and use (Sternberg, 1986). Philosophical theories tend to be competency based, psychological theories tend to be performance based, and educational theories employ both the psychological and philosophical approaches (Sternberg, 1986). Researchers note that the educational perspective to CT lacks the necessary clarity and rigorous testing of the theoretical frameworks when compared to the philosophical or psychological theories of CT (Ennis, 1985; Sternberg, 1986).

2.2.4 Critical Thinking and Domain Specificity

One significant unresolved theoretical issue among researchers is the domain specificity of critical thinking (Ennis, 1989). To some researchers, the teaching of CT skills occurs within a specific domain where background knowledge as essential (Case, 2005; Willingham, 2008). In this view, domain-specific knowledge includes understanding the principles and having the competence to engage in norm-regulated practices that make reasonable judgments in those specific contexts possible (Facione, 1990). Even further, Adler (1986) and McPeck (1981) argued that CT is always about some subject because students need something to think critically about (McPeck, 1990, p.10). Similarly, Bailin, Case, Coombs, and Daniels (1999) argued that domain-specific knowledge is necessary for CT because the kinds of explanations, evaluations, and evidence needed to make proper judgements vary among domains.

Ennis (1989) further articulated CT specificity by describing three types of specificity: Domain specificity, epistemological specificity, and conceptual specificity. Domain specificity is an empirically based view that CT requires 1) background knowledge of the subject, 2) is unlikely to transfer from one domain to another, and 3) is unlikely to be learned from general CT instruction (Ennis, 1989). The epistemological view of subject specificity posits that 1) background knowledge is essential for CT in a given field, 2) different fields hold different criteria for determining good reasoning, and 3) a full understanding of a field requires the ability to think critically in the field (Ennis, 1989). Conceptual subject specificity posits that CT does not exist outside of a specific subject matter and that the idea of a general critical thinking ability is meaningless, so general instruction in critical thinking is inconceivable (Ennis, 1989).

On the other hand, there are those researchers who conceptualize CT as domain general. For example, Van Gelder (2005) conceptualized CT as an intrinsically general construct applicable in a very wide range of domains that CT encompasses more than just thinking about some topic. In addition, Lipman (1988) alluded to the domain general nature of CT noting that CT facilitates good judgment because its fundamental meaning remains the same across the various domains.

While there is still much debate about the domain specificity of CT, researchers are beginning to acknowledge that CT has both domain specific and domain general aspects. For example, Ennis (1989, p.8) notes, that in math, deductive proof is the gold standard for reasoning whereas in the social sciences, statistical significance is most highly regarded, and in art, subjectivity is usually acceptable. However, Ennis (1989, p.8) acknowledged that there appears to be aspects of critical thinking that are common across disciplines, such as the notion that a conflict of interest detracts from the credibility of a source. In another example, Facione (2000) designed the California Critical Thinking Skills Test as a general test of CT rather than a specific test of CT. However, Facione (1990) underscored the importance of domain-specific knowledge in any application of CT skills and abilities. Thus, Facione (1990) acknowledged both the general and domain-specific elements of critical thinking.

2.2.5 Critical Thinking Transfer between Domains

Closely related to the domain specificity of CT skills is the extent to which CT skills and abilities transfer across domains. Halpern (2001) asserted that thinking skills must be applicable in a wide variety of contexts outside of the classroom otherwise they would not be beneficial to students. However, researchers note students may fail to transfer their CT skills and abilities across domains (Van Gelder, 2005; Willingham, 2008). The issue of CT transfer is especially challenging due to its very general nature and there are plenty of situations where CT skills can fail to transfer (Van Gelder, 2005, p.45).

One problem with discussing the transferability of CT skills is the ambiguity surrounding the “distance” of such transfer i.e., is the transfer “near” or “far” (Bailin, 2002; Ennis, 1989). For example, transfer may involve applying skills to a new but similar task, to an entirely new discipline, or from an academic context to a reak-world problem (Lai, 2011; McPeck, 1990). Accordingly, transfer of CT to tasks within the same domain is more likely to occur than to new disciplines (Lai, 2011, p.16).

2.2.6 Critical Thinking Disposition

Persons who are able to think critically possess the disposition to do so (Facione, 2000). They are able to use their CT skills appropriately, without prompting, and usually with conscious intent in a variety of settings (Halpern, 1998). The CT disposition describes the consistent internal motivation to engage problems and make decisions by using thinking (Facione, 2000, p.73). Furthermore, researchers show that challenging tasks emphasizing higher-order thinking skills may motivate students more than tasks emphasizing lower-order thinking skills (Turner, 1995).

Researchers generally agree that measures of CT should comprise both cognitive and dispositional components (Sosu, 2012, p.189). However, current measures of CT emphasize the cognitive aspects of CT more than its dispositional components (Halpern, 2003; Ku, 2009; Norris, 2003). Moreover, there is a dearth of instruments designed to measure the disposition to think critically (Sosu, 2012). This implies that the dispositional aspect of CT is, currently, not widely accepted (Halpern, 1998). The availability of suitable measures of the disposition toward CT is crucial to evaluating whether programs have been successful in nurturing CT attitudes in students (Ku, 2009).

In addition to the lack of emphasis on the dispositional component of CT and the paucity of instruments, researchers have debated the validity of the concept that the disposition toward CT is a separate construct to CT. Empirical evidence suggests that CT abilities and dispositions are, in fact, separate entities (Lai, 2011). Facione (2000) was able to identify seven critical thinking dispositions using the California Critical Thinking Disposition Inventory (CCTDI). Facione & Facione (1992) named the seven attributes of the disposition toward CT: Truth-seeking, open-mindedness, analyticity, systematicity, CT self-confidence, inquisitiveness, and maturity of judgment. Work by Facione, Giancarlo, Facione, and Gainen (1995, p.73) elaborate on these dispositions.

The truth-seeking disposition describes a person's eagerness to seek the best knowledge in a given context, is not afraid to ask questions, even if the findings contradicts their preconceived opinions and modify and evaluate their beliefs based on the information presented. Open-mindedness describes a person's tolerance to views different to their own and awareness of their personal biases. Analyticity describes process of solving problems and anticipating difficulties using reasoning and evidence. Systematicity describes a person use of an organized and diligent approach to solving problems or making decisions. CT self-confidence describes the amount of trust a person places in their own reasoning abilities. Inquisitiveness describes a person's intellectual curiosity especially when there is no clear application of information learned. Finally, maturity of judgment describes a person's use of good discernment when making decisions. When employing judiciousness in the decision making process there is an understanding that some problems have more than one plausible solution, and that many judgments are made using limited evidence.

2.2.7 Operationalizing Critical Thinking

Over the decades, several notable researchers have conceptualized CT in different ways. For example, John Dewey described CT in terms of reflective thought that involves the careful consideration of any belief given the evidence and possible conclusions derived from the evidence (Dewey, 1910). Later on, Edward Glaser (1941) described CT as comprising being thoughtful when presented with problems within the range of one's experiences, knowledgeable of the methods of logical inquiry and reasoning, and applying those methods. Following Dewey and Glaser, was Benjamin Bloom (1956) who described cognitive CT as comprising knowledge, comprehension, application, analysis, synthesis, and evaluation skills.

Stephen Brookfield (1987), conceptualized CT as consisting of emancipatory learning, dialectical thinking, and reflective learning. Emancipatory learning describes a learner's awareness of how situations influenced their thinking and taking the necessary action to change some aspect of these situations. Dialectical thinking describes a person's ability to understand contradictions and coming to a suitable resolution. Reflective learning involves a process of internal examination engendered by experience allowing person to come to a new understanding. Joanne Kurfiss (1988), director of the Teaching and Learning Center at Santa Clara University, defined CT as a process of exploring a task or problem to derive a hypothesis or conclusion that utilizes all available information, and can be convincingly justified. Diane Halpern, former president of the American Psychological Association, defines CT as thinking that is purposeful, reasoned, goal directed and uses a system of values to decide the most desirable outcome (Halpern, 2003).

John McPeck (1981) described CT as a context driven skill such that in the absence of context CT neither refers to nor denotes any particular skill. Richard Paul (1981) distinguished between 'weak' and 'strong' forms of CT because students' rely on their belief systems to process their experiences and find it easy to question information that they are not personally invested in but difficult to question information in which their personal investment is strong. In 1988, the American Philosophical Association (APA) funded a 2-year Delphi study to produce a common definition of CT and guide assessment. According the Delphi study (1988), CT is purposeful, self-regulatory judgments that result in the interpretation, analysis, evaluation, and inference, and explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which judgment is based (Facione, 1990, p.6). CT is non-linear because a person can apply the different components of CT to each other as well as to the problem at hand (Facione et al., 1995).

According to the Delphi study, interpretation defines a person's ability to comprehend and express the meaning of a wide variety of experiences by categorizing, decoding, and clarifying the meaning of information. Analysis defines the ability to discern both the intended and actual inferential relationships among statements and to examine and analyze the arguments presented. Evaluation requires an assessment of the credibility and plausibility a person's judgement. Inference requires the ability to identify the information needed to draw reasonable conclusions. Making inferences entails evaluating the evidence, formulating alternatives, and deriving conclusions from the data. Explanation is a cognitive skill involving being articulate when describing the reasoning underlying a particular response. Self-regulation involves engaging in the process of metacognition and using that information to evaluate one's own inferential judgments with the aim of correcting the person's reasoning or results.

Given the plethora of CT definitions provided, this study adopted the definition of CT provided by Kurfiss (1988) that described CT as "an investigation whose purpose is to explore a situation, phenomenon, question, or problem to arrive at a hypothesis or conclusion that integrates all available information, and can be convincingly justified" (p.20). This definition of CT encapsulates sustainability and the challenges of creating a sustainable world. As indicated earlier, the entire concept sustainability is a challenge that encompasses a myriad of issues that require an ability to thoroughly explore all the facets of sustainability and integrate these components in a cohesive whole. This alludes to the CT skill of exploring a situation and integrating information outlined in the definition. Moreover, as noted earlier in the literature review achieving sustainability requires fundamental social changes that would challenge advocates of sustainability to justify in a convincing manner why these changes need to occur.

Overall, HEIs and organizations are working towards creating scales to measure students' knowledge of sustainability. Unfortunately, a lack of consensus on the definition of sustainability hampers their efforts to measure and evaluate sustainability. Additionally, the studies provide little or no evidence of reliability and validity and focus largely on the ecological aspect of sustainability. Similarly, researchers and HEIs have also worked to measure students' ability to think critically. However, many of these CT scales measure critical thinking in fields such as mathematics and reading, but not sustainability. Moreover, there is notable lack of studies that measure undergraduate students' ability to think critically about sustainability. Given the importance of both sustainability and CT to students' educational and social development, there is a need for a reliable and valid measure of students' ability to think critically about sustainability. Appendix B provides a summary of critical thinking scales found in the literature.

CHAPTER 3. METHODS

The primary objective of this study was to develop a reliable and valid scale to measure undergraduate students' ability to think critically about sustainability named the Critical Thinking about Sustainability Scale (CTSS). This study utilized a sequential mixed methods research design that consisted of a qualitative phase followed by a quantitative phase to address the following two research questions (RQ): 1) How do sustainability experts vary in their conceptualizations of sustainability? 2) To what extent does the CTSS measure students' ability to think critically about the ecological, economic, and social aspects of sustainability?

The main purpose of the qualitative phase was to answer the first research question. The data collected from the qualitative phase aimed to help specify a definition of sustainability by integrating expert's conceptualization with the literature. The main purpose of the subsequent quantitative phase was to answer the second research question. This part of the study involved the development and administration of the CTSS, the evaluation of the construct validity of the scale and the evaluation of the quality of the questions on the scale.

3.1 Exploration of Experts' Conceptualization of Sustainability

Subsequent to receiving Institutional Review Board (IRB) approval, this study recruited sustainability experts via the snowball sampling method. In this non-probability sampling method, faculty members who are experts in sustainability identified other sustainability experts among their colleagues who may be willing to participate in the study. In Fall 2014, an email was sent to prospective sustainability experts to inquire about their willingness to be interviewed for the study. In total, five experts agreed to participate in the interviews. Expert A is a professor in the engineering education department and has 16 years of experience teaching and doing research at the university level. Expert A has engaged in research and publication about sustainability in engineering. Expert B is a professor with a joint appointment in civil engineering and environmental ecological engineering and has 21 years of experience teaching and conducting research at the university level. Expert B teaches undergraduate and graduate courses related to sustainability.

Expert C is a professor in the School of Engineering Education, and is an affiliate faculty in the Women's Gender and Sexuality Studies and the Division of Environmental and Ecological Engineering. Expert C has 10 years of experience teaching and doing research at the university level. Expert C has also co-authored books, journal articles related to sustainability, and has received an external grant to support research about sustainability in engineering. Expert D works in the Office of University Sustainability, and has 3 years of experience teaching and doing research at the university level and 16 years of experience in the architecture industry with a focus on green-architecture. Expert E works at the Discovery Park Center for the Environment and has fourteen years of experience teaching and doing research at the university level.

Together, the sustainability experts with a diverse range of experiences both in academia and industry related to sustainability were able to give valuable insights about the topic. I developed a brief interview protocol to guide the interview process (Appendix C). The protocol consists of four key questions, and the interviews lasted between 30-60 minutes in length. Prior to each interview, I asked the interviewees for their permission to record the interview. I used Inqscribe software (Inquirium, 2005) to transcribe and de-identify the data, thus no personally identifiable information is retained in the transcripts. I use the coding procedures suggested by Stauss and Corbin (1990) to analyze the interview data. I began with an open coding procedure for data analysis in order to identify categories and sub-categories in the data.

Next, I used axial coding to systematically develop the categories and link these categories to the sub-categories. Then finally, I used selective coding in order to integrate and refine categories. After defining the categories, I evaluated the similarities and differences in experts' conceptualization of sustainability with those derived from the literature. Because I asked four specific questions related to sustainability to each participant at the interview, I analyzed the expert's responses to each question separately using the qualitative data analysis software NVivo 10 (QSR International, 2014). The use of the Nvivo software allowed for the persistent observation of the data to identify those elements of the data that are most relevant to the research questions and focusing on those elements in detail to provide depth to the quality of information derived from the data (Lincoln & Guba, 1985). The use of the qualitative data analysis software also allowed for better triangulation of the data with the literature to get a more complete understanding of the experts' responses (Lincoln & Guba, 1985). Therefore, the use of the Nvivo software helped add a level of trustworthiness to the data.

3.2 Scale Format Selection

In this study, I used scenarios to measure undergraduate engineering students' ability to think critically about sustainability. Since the 1950s, researchers have utilized scenarios in engineering (e.g., Chemical and Civil engineering) as a means of teaching and motivating students by placing them in authentic contexts involving problem solving or making decisions (Prince & Felder, 2006; Yadav, Subedi, Lundeberg, & Bunting, 2011). Moreover, research has shown that scenarios help promote students' conceptual understanding and higher-order thinking skills (Mayo, 2002; Yadav, et al., 2007). Therefore, given these considerations I utilized scenarios for the development of the CTSS.

The scenarios I used in this study were sourced from the Living on Earth (LOE) website (loe.org) and the National Center for Case Study Teaching in Science (NCCSTS). With the guidance of Expert B, I selected six scenarios for further evaluation: four scenarios came from the LOE website that allowed free access to the scenarios and two from the NCCSTS website for which I had to obtain permission to use (see, Appendix D). Selection criteria for the scenarios were relevance to students and the novelty of the topic. Relevance to students refers to how well students of diverse academic backgrounds can understand the scenario. Novelty of the topic refers to how interesting the topic would be to students. I selected three of the six scenarios for the development of CTSS. The first scenario called "Keeping Foods and the Climate Cool (KFCC)" looked at food waste in developed and developing countries (Appendix E). The second scenario named "Broken Bulb Dreams (BBD)" involved the use of incandescent light bulbs and their potentially harmful effects on the environment (Appendix F). The third scenario called Ode to Ogallala (OTO) focused on the use of fossil water from the Ogallala aquifer for agriculture and industry (Appendix G).

Each scenario is approximately 1-page in length (double-spaced) followed by seven open-ended questions. I use open-ended questions for four major reasons. First, and most importantly, open-ended questions require students to utilize complex thinking skills (Badger, 1992). Second, open-ended questions can generate rich, detailed answers that can provide valuable information for understanding respondents' thinking (Davino, 2013). Third, open-ended questions address the key concepts, processes, and skills that go beyond specific instructions. Fourth, open-ended questions allow students to solve problems in a natural way. The advantages gained by the use of open-ended items are ideal for measuring critical thinking skills (Ennis, 1993; Halpern, 2003; Renaud & Murray, 2004)

However, for all their positives, open-ended questions are often criticized for their difficulty of controlling measurement errors occurred through coding and analysis processes (Davino, 2013) and if poorly designed for the response task, they may increase respondents' chances of making errors like providing verbal explanations where numeric responses were required (Christian, Dillman, & Smyth, 2007). Despite these drawbacks, open-ended questions are the most appropriate medium for investigating engineering students' ability to think critically about sustainability in this study for the advantages indicated earlier. In addition, given the fact that the CTSS is in the early stages of scale development, I also aimed to evaluate the open-ended questions for the potential presence of these drawbacks and their effects on the validity of CTSS.

3.3 Item Development and Scoring

There were seven questions in total. The aim of Questions 1 to 6 was to allow respondents to demonstrate their ability to think critically about sustainability, and the aim of Question 7 was to allow students to express their ideas about what sustainability meant to them. The purpose of Question, 1was to examine students' ability to discern the three aspects of sustainability given the information in the text by asking them to describe the sustainability issues in the text. Question 2 aimed to evaluate students' ability to identify missing information and incorporate the information to the context of the scenario. Question 3 aimed to measure students' problem solving ability given limited information by asking students to recommend a strategy for resolving the sustainability issue and examining how or why the strategy would resolve the issue. Question 4 assessed students' abilities in identifying, analyzing, and evaluating the different positions of the stakeholders involved and the alignment of their interests. Question 5 was designed to examine students' evaluative skills by asking them to describe the possible implications or consequences of their recommendation(s) provided in Question 3. The goal of Question 6 was to evaluate students' metacognitive skills by asking them to describe how their own knowledge, perspectives, and opinions influenced their interpretation of the case study. Finally, the goal of question 7 was to understand students' perceptions of sustainability by asking them to describe what the term sustainability meant to them.

I developed Questions 1, 6, and 7 mainly by referring the qualitative results on experts' conceptualization of sustainability and students' knowledge of sustainability. I developed Questions 2, 3, 4, and 5 mainly, but not exclusively, using knowledge also obtained from the qualitative results on students' knowledge of sustainability, measuring students' knowledge of and ability to think critically about sustainability.

To evaluate students' responses to the questions, I created a scoring rubric based on the Purdue University's Core Curriculum, Learning Outcome Rubrics (Appendix H). Expert B reviewed the draft rubric and suggested edits to the rubric concerning the wording of the categories. Students received scores based on a three-point rating scale indicating: proficient = 3 points, emerging = 2 points, and developing = 1 point. The scoring rubric was developed such that student responses to the questions on each scenario could be accurately scored regardless of the content of the scenarios. The rubric provided a fair and systematic framework to ensure that the scoring of the questions on the scenarios was as fair as possible. The criteria for receiving a rating of proficient, emerging, or developing are uniquely developed for each question. For example, Question 1 asked students to describe the sustainability issues and challenges explaining the pros and cons of these challenges. A student would obtain a score of 3 points (proficient) if they fully described the economic, social, and environmental issues of the scenario. If the student fully described two or three of the sustainability issues, they would get a score of 2 points. If the student only described one sustainability issue, they would receive a score of one point. Note that the scoring procedure resulted in data that are ordinal in nature and hence do not follow normal distributions, thus necessitating the use of statistical methods capable of analyzing ordered-categorical data. Finally, I excluded Question 7 from the scoring criteria since the goal of question seven was to determine what the term "sustainability" meant to students and not to evaluate the quality of their responses.

3.4 Evaluation of the Content Validity of the CTSS

To establish the content validity of the CTSS, I submitted the scale for review by a group of content experts and a group of students from the sample population. Three persons served as content experts for the evaluation of the CTSS: sustainability Expert B, an engineering faculty knowledgeable about sustainability and recommended by Expert B, and a graduate teaching assistant for an undergraduate course about sustainability. The student reviewers were three undergraduate teacher's assistants (UTAs) to a sustainability course.

The undergraduate student TAs provided valuable information concerning the relevance of the topic to students from diverse academic backgrounds and the novelty of the topic to students. This information helped made the CTSS more amiable to the potential participating sample of students (i.e., undergraduate engineering students) in the current study. The content experts and the UTAs reviewed the scenarios and questions over the course of one week and gave their recommendations for improving the scenarios regarding content, structure and wording of the text, as well as the quality and ordering of questions. For example, a majority of the reviewers recommended adding more information to each of the scenarios to provide greater context for the participants, reordering and rewording of some of the questions and the initial scale was modified based on the reviewers' comments and reviewed by Expert B a second time.

3.5 Evaluation of Construct Validity

The purpose of this phase of the study was to appraise the questions designed to measure students' ability to think critically about sustainability using a series of psychometric analyses. This section included information about the sample and sampling procedures for obtaining sufficient sample size to conduct the quantitative analyses, scale administration, and evaluation of the scale dimensionality and quality. The data analysis for establishing construct validity occurred in two sub-phases: the first of which involved assessing the dimensionality of the test using exploratory factor analysis (EFA) and scale quality analyses based on item response theory (IRT). EFA was chosen over confirmatory factor analysis (CFA) because this study is in the early stages of the CTSS development and no specific theories about the nature of the data are indicated (Brown, 2006). This is in contrast to CFA where the dimensionality of the data is established and there is a need for a parsimonious model to test specific theories (Brown, 2006). This dimensionality evaluation also served as a necessary selection tool for the most appropriate item response model.

Note that I evaluated the data for each scenario separately because there was a possibility that the results of the EFA and IRT analyses would necessitate the replacement of a scenario(s) due to poor dimensionality or item quality. Moreover, one should exercise caution when pooling data from samples that are known to be different on some criteria (in this case different scenarios) or are repeated measurements (as is in this study) for factor analytic purposes (Tabachnick & Fidell, 2007). This is because, within the context of this study, the three scenarios may produce different factors or students' responses to the questions may improve or worsen across scenarios, therefore pooling the data may obscure differences among scenarios rather than reveal them (Tabachnick & Fidell, 2007).

3.5.1 Sample and Sampling Procedure

The target sample for the pilot testing of the developed scale were all levels of undergraduate engineering students 18 years or older at a large Mid-Western University during the Spring 2015 and Summer 2015 semesters. The target sample pool consists of approximately 7,877 students, of whom 23.1% are female, 76.9% are male, and 6.4% of the sample is underrepresented minorities. The sampling frame does not include graduate students because the vast majority of research on sustainability focuses on undergraduate students so that the developed scale intends to use for undergraduate students. Additionally, I placed no restrictions on the undergraduate engineering students who could have participated in the study. After acquiring IRB approval, I contacted several Engineering faculty who were sustainability experts regarding their possible participation in my study. The faculty in turn recommended other sustainability faculty that I could contact. Through this purposive sampling design, I was able to enlist the participation of two engineering faculty for the study. I provided the participating instructors with a copy of the IRB information sheet containing the details of the study. I presented a brief description of the study to one of the participating classes, while the teacher's assistant (TA) of the second participating class provided information about the study to the students. In both instances, the students knew that their participation was voluntary, that they may skip any questions they may not wish to answer, and their responses will be anonymous.

In addition to a verbal description of the study, the students also received an electronic version of the IRB information sheet along with the three scenarios. With the permission from the students and the instructor, I coordinated with participating faculty and the custodian of the student list-serve (in the case on one faculty member) to administer the cases to the participants. The students received extra credits for their participation after discussion with their faculty and/or listserv custodian. A total of 209 students responded to the questions on the CTSS. After the data collection, I excluded responses from students who provided no responses to any of the scenarios, who provided responses to only one or two of the three scenarios, students who provided duplicate responses, students who did not respond to any of the questions on the scenarios. The final valid sample consisted of the 151 responses with complete data on all questions for all scenarios corresponding to a response rate of 72%. Table 3.1 shows the demographic characteristics of the student sample based on 151 cases.

Table 3.1
Demographic Characteristics of Participant Sample (N = 151)

	N	Percent (%)
Gender		
Male	102	66.9%
Female	49	32.5%
Class		
Freshman	26	17.2%
Sophomore	72	47.7%
Junior	31	20.5%
Senior	22	14.6%
Major		
Aeronautical Engineering	1	0.7%
Agriculture Engineering	1	0.7%
Atmospheric Science	1	0.7%
Biological Engineering	3	1.9%
Chemical Engineering	7	4.6%
Civil Engineering	11	7.3%

3.5.2 Scale Administration

I administered the scale to the students using the Qualtrics online survey tool. Administering the scenarios online was desirable for this study because it allowed me to distribute the scenarios quickly to a large number of students at very low cost and in a short period of time (Dillman, Smyth, & Christian, 2009). In order to obtain a deep understanding about students' ability to think critically about sustainability it is vital to obtain accurate records of students' responses. The online administration of the scenarios helped with this by allowing for immediate data collection, the ability to track student progress on the survey, and allowing students to submit their responses to the database that they could easily edit their responses if needed. For one participating group of students, I sent the link to the scenarios to the custodian of the student listserv who then distributed the link to the students. For the other participating group of students, I sent the links to the instructor who then administered the scenarios to the students. The students had two weeks to complete the scenarios. For the first group of students the survey was available on April 24th to May 7th, 2015. For the second group of students, the survey was available on June 17th to July 1st, 2015. I sent reminder emails approximately three days prior to the closing of the survey and I granted additional time (at most 48 hours), on request of the participating faculty, to students who required an extension to complete the scenarios.

3.5.3 Evaluation of CTSS Dimensionality

As previously mentioned, I conducted an EFA for ordered-categorical data to assess the dimensionality of the construct and by extension the content validity of the scenarios. As a first step, I evaluated the reliability of the items on the scale by computing Chronbach's Alpha (α) for each scenario, the item-total correlation (ITC) and the α-if-item deleted statistics. For the purposes of this study I consider $\alpha = 0.70$ as adequate (Nunnally, 1978). Given the purpose of the CTSS and the process of its development, I assumed that the scenarios of the CTSS would measure only one construct: “Students’ ability to think critically about sustainability”.

Due to the nature of the data (i.e., ordinal), I evaluated the polychoric correlation matrix to identify patterns of inter-item correlations that may give a preliminary appraisal of the underlying structure of the data to justify the use of EFA. More specifically, I assured at least several correlations are greater than 0.3 because a correlation matrix that is amenable to factor analytic procedures should include several correlations above 0.30 and if no correlations exceed 0.30 then the use of EFA is questionable (Tabachnick & Fidell, 2007).

To judge the number of factors to extract I examined the scree plot to identify the point where the discontinuity in the eigenvalues occurs and retained the factors on the curve before the first point that starts the discontinuity. As an added measure, I also evaluated the size of the eigenvalues that are a measure of variance. From a variance perspective, a latent variable with eigenvalue less than 1 is not as important as one with an eigenvalue greater than one (Tabachnick & Fidell, 2007). Therefore, I selected the number of factors to extract based on the number eigenvalues greater than one.

Given the fact that I assumed the CTSS measured one construct I utilized the default oblique rotation method in Mplus 7.1. To obtain a simple factor solution I used the following selection criteria, I retained factor loadings above the 0.32 on a single question, deleted questions with loadings lower than 0.32, deleted negative factor loadings from the scale, deleted questions with loadings greater than 0.32 on more than factor (Brown, 2006). Factor loadings greater than 0.50 were considered strong (Costello & Osborne, 2005) and all loadings were evaluated at alpha = 0.05.

To evaluate the fit of the model to the data I utilized a number of fit indices obtained with the WLSMV estimator when specifying the EFA model in Mplus 7.1. I evaluated the fit of the EFA model to the data using the following fit indices: a Chi-square statistic, Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993), Comparative Fit Index (CFI; Bentler, 1990), Tucker Lewis Index (TLI; Tucker & Lewis, 1973), the Standardized Root Mean Square Residual (SRMR; Bentler, 1995; Jöreskog & Sörbom, 1986), and for negative residual variances. The Chi-square statistic evaluates the fit between the sample covariance matrix and the estimated covariance matrix, and a non-significant result indicates a good fit (Tabachnick & Fidell, 2007). The RMSEA measures the lack of fit between the estimated model to a perfect model, where values < .05 are excellent, and values < .08 are good (Brown & Cudeck, 1993). The CFI measures the relative fit of the estimated model compared to a baseline model where values > .95 are excellent (Hu & Bentler, 1999). The TLI provides an index of increment of fit comparing a null model of independence (i.e., no factors) with a model with the requisite number of factors (Tucker & Lewis, 1973).

The SRMR is a measure of the mean absolute correlation residual and is the overall difference between the observed and predicted correlation matrices; and for the SRMR values below 0.08 indicate good fit (Hu & Bentler, 1999). Previous studies have shown that these indices function well for detecting model misspecification and are not overly dependent on sample size (Hu & Bentler, 1998, 1999; Marsh, Hau, Balla, & Grayson, 1998). I also investigated the presence of negative residual variances that tend to indicate model specification errors, identification problems, or input errors (Bagozzi & Yi, 1988).

3.5.4 Evaluation of CTSS Question Quality

The second phase of the analysis involved evaluating the quality of the questions on the scale using appropriate IRT methods. Frequently utilized IRT models (i.e., a family of unidimensional IRT models) require a single continuous latent variable that represents individual differences on the underlying construct. The degree to which the data meet the unidimensional assumption affects the interpretation of item and scale information functions, the estimation of individual trait levels, the evaluation of cross-group differences of differential item functioning, and/or the use of tests such as computer adaptive testing, and linking scales (Reise, Cook, & Moore, 2015). The fitting of a unidimensional IRT model to multidimensional data or vice versa can distort parameter estimates (Steinberg & Thissen, 1996) and undermine the validity of obtained coefficients (Reise, Scheines, Widaman, & Haviland, 2013).

Therefore, the dimensionality analysis in the preceding section served as a necessary tool to determine the most suitable IRT model for evaluating the quality of the questions. After establishing the unidimensionality of the construct via EFA, I utilized the Mplus 7.1 software to fit a unidimensional 2-paramter graded-response model (2PL GRM) using the maximum likelihood estimator with robust standard errors (MLR) which is robust to non-normal data structures, to the data to determine the quality of the questions on the scale. The 2PL GRM is appropriate because the data are ordered-categorical in nature. In addition, there is a concern about dependency of questions within each scenario because the questions in each scenario all relate to a common stimulus. As a result, the important IRT assumption of local independence may be violated leading to biased estimates and inaccurate inferences about the data. However, researchers have found that the graded response model is likely to robust to the violation of the local independence assumption (Cook, Dodd, & Fitzpatrick, 1999). Additionally, the IRT methods used to analyze the data accounts for measurement error, thus providing an added layer of validity to the parameter estimates (Muthén & Muthén, 1998-2012). This reduces the potentially negative impact of local dependence thus increasing the confidence in the inferences made from the data.

I opted to not use the 3-paramter graded-response model (3PL GRM) because research has shown that the pseudo-guessing parameter in the 3PL GRM may not be well estimated because of a lack of information at the low end of the ability scale (Lord, 1980; Reckase, 2009). This poorly estimated guessing parameter may affect the estimation of the other item parameters (Swaminathan & Gifford, 1985). Researchers have also shown that the 3PL GRM does not necessarily provide a better fit than the 2PL GRM (Yen, 1981). Additionally, studies have indicated the gain in the data description from the 3PL GRM is small given the computational difficulty associated with the estimation of the guessing parameter (Haberman, 2006). Given these considerations and the fact that the study is in the early stage of the CTSS development, I utilized the 2PL GRM instead of the 3PL GRM.

The 2PL GRM describes questions in terms of slopes and thresholds. The slope parameter represents the questions' ability to distinguish between examinees with different ability levels. Generally, questions with higher slope parameters provide more information (Hays, Morales, & Reise, 2000). Threshold parameters represent the trait level necessary to respond above threshold with 0.50 probability (Hays et al., 2000). This parameter characterizes the boundary between response categories and is equal to the number of categories minus one. Because I used a 3-point rating scale (1, 2 and 3) in this study, the number of threshold parameters is 2.

Next, I evaluated the quality of the questions by assessing the item information functions (IIFs), and the test information functions (TIFs) for each scenario. In the context of IRT, information represents a statistical indicator of the quality of the estimate of a parameter at a particular trait level (Reckase, 2009). Typically, item information are presented as functions of the parameters under investigation rather than a single value because item responses provide information at all levels of values on the scale (Reckase, 2009). Moreover, different items can provide different amount of information among varying levels of a latent trait (Reise, Ainsworth, & Haviland, 2005). Therefore, the item is most informative at the trait level where the item information function peaks and as one moves away from the peak, the question provides less information indicating that the trait scores are not estimated very well (Thorpe & Favia, 2012).

Additionally, because information are directly related to measurement precision, as one moves away from the peak of the information curve, the precision (reliability) of the scores on ability decreases (Baker, 2001). In the case of the graded response model, the information curve may show the most information at more than one 'peak' and this corresponds to the number of thresholds in the estimated model and this may cause problems for the selection of appropriate questions for scale design (Thorpe & Favia, 2012). Because the model adopted for this study involves two thresholds, it is possible to have two peaks where the trait level provides the most information.

The following criteria suggested by Baker (2001) were used to evaluate the magnitude of the slope parameter: none = 0, very low = 0.01 – 0.34, low = 0.35 - 0.64, moderate = 0.65 – 1.34, high = 1.35 -1.69, very high > 1.70. Because information is proportional to measurement precision (more in formation equals more precise measurement), the test information curve estimates how well a test functions as a whole in different trait ranges (Reise, et al., 2005). For this study, I evaluated the information based on the shape of the item information curves namely how peaked or flat they are which is an indicator of their tendency to differentiate among students of varying ability along the continuum and whose thresholds capture a range of the latent construct. I evaluated the fit of the model to the data using the Chi-square statistic that indicate the agreement between the observed proportions of correct responses and the responses obtained by the fitted model for an item where a non-significant Chi-square statistic indicated a good item-fit (Baker, 2001; Embretson & Reise, 2000).

Overall, the methods described in this section ensure to address the research questions of the study by enabling the efficient evaluation of the dimensionality of the CTSS including important information concerning the fit of the model to the data and a means of reliably determining the quality of the questions on the CTSS.

CHAPTER 4. RESULTS

This chapter presents the results corresponding to each of the research questions regarding the development and validation of the CTSS. The section begins with a presentation of the results from the qualitative phase that provides answers to research RQ1: “To what extent is sustainability experts’ conceptualization of sustainability consistent or different among each other?” The answer to this question came from the results of the one-on-one interviews with sustainability experts. The results of the qualitative phase also served to help specify an operational definition of the term “sustainability” that was derived from both the interviews, and the literature gathered on sustainability.

Next, I present the results of the quantitative phase of the study. The results of this phase answered RQ2: “To what extent does the CTSS measure students’ ability to think critically about the ecological, economic, and social aspects of sustainability?” This question addressed the construct validity of the CTSS using EFA (presented first) and the quality of the questions of the CTSS (presented second) using the 2PL GRM. The section ends with the summary of the findings.

4.1 Experts’ Conceptualization of Sustainability

The results of this section describe how sustainability experts were similar or different in their responses to four questions concerning sustainability during the one-to-one interviews (RQ 1). These interview questions were: 1) What does the term sustainability mean to you? 2) In your opinion, is there difference between sustainability and sustainable development? 3) In your experience, what do students know about sustainability? 4) How would you or how have you measured students' knowledge of sustainability? The results follow in the order of the questions in the interview protocol. Additionally, at the end of the section, one will find an operational definition of sustainability within the context of the study and Table 4.1 that depicts a brief summary of the similarities and differences of experts' conceptualization on the different aspects of sustainability.

4.1.1 Variations in experts' conceptualization of sustainability

One of the key ideas emerging from this question was that sustainability was a difficult concept to define and that there was no one right definition of sustainability. For example, Expert A explained:

I would say the definition (sustainability) is something like using the resources today without impairing the future generations' ability to live and thrive. That really we are looking at the big picture, we are looking at all situations, and we are looking far into the future as well as today.

While Expert B conceptualized sustainability in a different manner:

It's all about context…the meaning of it (sustainability) is determined by the context of the question, the persons, and the field. In environmental engineering, there is an understanding that we are talking about environmental and ecological sustainability. Businesses use the word all the time to talk about the sustainability of their business model, and without context, it is almost a meaningless word. So, but big picture sustainability is about continuing, sustaining maintaining and in the context of environmental and ecological engineering sustainability is about sustaining human civilization and ecosystems that support human civilizations in the foreseeable future, millennia even.

For Expert C, sustainability begins and ends with concerns about protecting the environment. As Expert C explained:

I think currently what sustainability is having to mean is sort of trying to…look out to the future and anticipate what environmental problems are going to be in the future. So, I know there's the whole environmental, social, economic sustainability stuff or the people, prosperity, planet sustainability definition. But for me, at the end of the day it's all about environmental issues, but part of the complexity of the term has to the do with the fact that we don't know what impact we are going to have on the planet.

When compared to the other three experts, Expert D has a different notion of sustainability:

I think it's a challenge to find out what it is (sustainability)...So you know there's lots of different definitions…the definition (sustainability) I am liking of late is, 'sustainability is the ability of an organism to remain diverse and productive over time. The organism can be a community, a company, higher ed, or a planet. I like it because it addresses the time component.

Expert D went to note:

A lot of us who practice sustainability are heavily in the environmental lobe of the triple bottom line…but I think you have to address all three, that if you only get one or two then you are kind of missing the full picture. There is a social justice and people component. I think if you look at the different systems that address sustainability performance the environment is well addressed …the economic is up there too, the social and human element is probably the least defined.

Therefore, it is clear that the experts all viewed sustainability from different contexts and that the term was difficult to define. Another common theme that permeated the experts' conceptualizations was the temporal aspect in sustainability. They all described the concept of time—whether it was about the "future" or "over time", the experts all included a time component to their description of sustainability. The idea of equity was another theme that came up in experts' conceptualization of sustainability—though not always explicitly. The explicit mention of equity was dependent on how they conceptualized sustainability. This is clearly demonstrated by Expert C who said "I have a book about social justice in engineering, so it's not that I think that the concept of social sustainability is unimportant. I just do not think they need to be hooked to the wagon of sustainability".

Although they shared the idea of sustainability having economic, ecological, and social components, the experts differed in how much importance they attributed to the various components of sustainability. For example, Expert C situated their conceptualization of sustainability firmly within ecological perspectives in contrast to Expert D who advocated the acknowledgement of all three components of sustainability.

There were other unique perspective reveled by the experts. For example, Expert C pointed out that people tended to describe the environment using feminist terms such as "mother earth" or "mother nature". Therefore, to Expert C, sustainability also engaged discussions about the male-female dichotomy. While Expert A framed sustainability within the context of making decisions (engineering) that are viable to humanity both the present and the future and even described a political aspect to sustainability explaining that students' political stance may affect the teaching of sustainability in the classroom. Expert A notes that:

Sustainability is always in the background, in remembering that students come in with a political persuasion. They are probably predisposed somehow to at least the language that is used, and again I mentioned before the word 'sustainable' itself in some aspects of the US political life it is a bad word. It (sustainability) is evil, and it is trying to take over and there are people who believe that and that is reasonable, I suppose, for them to believe.

This expert went to on to say:

As educators, we need to be sensitive to the idea that there are people who think like that (sustainability is bad) and therefore say 'okay when we use the word sustainability what we're talking about is broad values. The values of long term planning, of understanding the system, and not necessarily the values that the UN is putting forth, because we know that some of you don't like the UN.

Overall, one can see that consistent with previous literature, the experts were consistent in their conceptualization of sustainability as an idea that is difficult to define. These results seem to lend support to the idea, espoused most clearly by Expert B, that in order to understand sustainability there is need to understand the context from which the idea of sustainability is being discussed. All experts agreed sustainability consisted of a time component whether they utilized the word "future" outcomes are considered in present choices, and where equity within and between generations are maintained. All the experts acknowledged that sustainability consisted of an economic, ecological, and social aspect but they differed in the emphasis they placed on the aspects of sustainability. For Expert C, sustainability was all about the environment to the extent that the other aspects of sustainability were dependent on the achievement ecological sustainability. Whereas, for Expert D, all aspects of sustainability are important and to choose one aspect of sustainability over the other would mean to lose sight of the other aspects of sustainability that are vitally important.

4.1.2 Sustainability vs. sustainable development

Regarding the experts' views concerning sustainability vs. sustainable development there were variations in how experts answered the question. There were two general views expressed by the experts on this question concerning sustainable development; 1) development as an improvement of quality, and 2) development as a subset of sustainability.

Experts A and B viewed sustainable development as the improvement of quality or making something better. For example, Expert B explained that the differences between sustainability and sustainable development depend upon the definition of the term "development" as qualitative (e.g., improving education) or quantitative (e.g., constructing buildings, clearing land). In this expert's view, from a qualitative perspective, sustainability and sustainable development can mean the same thing (i.e., the improvement of quality of human life). Therefore, if development does not lead to an improvement in the quality of life then sustainability and sustainable development are contradictory since in Expert B's view sustainability is all about improving the quality of life. Expert B also mentioned that sustainability without development is not necessarily fair to everyone around the world.

Experts C and D viewed sustainable development as a subset of sustainability. Expert C viewed sustainable development as a tool designed by large western countries to absolve themselves for feeling guilt for benefitting from industrial development that smaller nations lacked, and which have had serious negative impacts on the environment. Thus, sustainable development is a nice tool used by industrialized countries to help underdeveloped (non-industrialized) countries advance in a more sustainable way than larger countries did. Expert C notes:

I think that development is kind of a weasel idea. I think that there is a global inequity that countries who had successfully industrialized and raised the quality of life for the majority of its citizens, did so on the backs of countries that did not. So that countries that are looking to industrialize, are doing so in world where climate change is happening at a profound rate. I think that sustainable development is a term that western people use to absolve themselves of feeling guilt that they have benefited from industrialization without regard to the impact of climate change which developing countries can no longer do...they don’t get that privilege.

Expert D viewed sustainability as encompassing everything such as water, wastes, transportation, energy, and food systems, while sustainable development involves how to achieve human needs in a more sustainable way. Thus, in this expert’s view sustainable development is a subset of sustainability.

4.1.3 Students' knowledge of sustainability

The consistent view regarding students' knowledge of sustainability among the experts is that students have limited knowledge about sustainability. Generally, their students tend to equate sustainability with the environment, pollution and recycling. Their understanding of sustainability was one-dimensional and abstract. Even with this limited knowledge, the students do not know the details about the environmental issues. Expert D noted qualitative differences between graduate students and undergraduate students in their attitudes towards sustainability. Expert D notes:

I will say that in terms in doing projects, real projects, we had more luck with graduate students than with undergraduate students. The graduate students are more committed they are willing to say: "It's a year-long project I'm in." The undergrads are more like: "I've got to take this class" and as it gets closer to the end it's sort of "I've got to juggle all these things and this is not really that important". In talking with my peers that is pretty consistent across the country.

Expert C explained that students tended to have a defined perception of what an engineering project looks like—they understood the concepts of making a tool or project more efficient or including alternative energy sources. However, students did not consider the sustainability issues that underlie the engineering project or the implications of the project on the society. The expert went on to note that students have no framework to think about their ethical and moral obligation to the community or nation as opposed to individual experiences. Students tended to think that the bigger environmental story is irrelevant to them. Expert C notes:

However what does exist tends to be about micro ethics it’s about not lying cheating, stealing, being a good engineer and signing off only on stuff that you are able to asses as good. It is not about macro-ethics, like "What is my role in the production of consumer goods in this country that then has to be disposed of, it uses energy, it uses energy in the production, it pollutes?

Another important theme emerged in this question is potential gender differences in knowledge in sustainability among undergraduate students. For example, Expert B noted:

Once in while we will have one or three students who will not engage; they will just not engage, and will be extremely resistant to engaging to the point of being disruptive in class. We push them into the affective domain "how do you feel about this?", "what is your opinion about this?" They then become disconcerted, uncomfortable and will be disruptive—these students are always men.

Expert B also noted that women are more reflective and willing to make sacrifices for others to see the need to do right for all people. This is a noteworthy revelation because science technology engineering and math (STEM) fields tended to be male dominated (Ginther & Kahn, 2006). Expert Bs revelation coupled with Expert Cs description of sustainability as feminine concept highlights a potentially key idea about sustainability that does not seem to be addressed to great extent in the literature.

4.1.4 Measuring students’ knowledge of sustainability

A common theme occurring in experts’ responses to this question was that critical thinking is key for understanding sustainability. All the experts stressed that critical thinking was a key aspect for understanding sustainability and that assessments should reflect their ability to think critically about sustainability. However, although all the experts mentioned critical thinking as crucial to understanding sustainability their attempts to measure critical thinking in their students has been less than optimal. Generally, all the experts tended to measure students' knowledge of sustainability using tests and quizzes. When measuring students' ability to think critically about sustainability projects were the main assessment tool.

For example, Expert C focuses on the basic meaning and/or definitions of the various sustainability terms i.e., improving their content knowledge and they measure how well students are able to apply their content knowledge to their design projects. Expert C would have their students review their peer's designs as well as reflect on their own work. However, Expert C saw limited use of critical thinking skills in those areas. However, the expert noted that the critical thinking aspect of assignments are competing with a lot of stuff other students are thinking about. The experts all shared the view that there are limited resources that measure students' ability to think critically about sustainability and more work need to address this notable gap in engineering students' educational development.

Table 4.1 below provides a summary of the similarities among the sustainability experts in their responses to the questions as well as unique responses from the interviews.

Table 4.1
Similarities and Differences in Sustainability Experts' Conceptualization of Sustainability Issues

Sustainability conceptualization	
Similarities	Unique ideas
• Experts' sustainability definition included a time component • Experts described sustainability as a difficult idea to conceptualize • Experts mentioned equity in the distribution of the earth's natural resources • Experts mentioned equity in opportunities for people and societies thrive and survive	• Experts tended to emphasize one aspect of sustainability more than other • Two experts described sustainability as a feminine construct • Expert A described a political aspect to sustainability
Sustainability and Sustainable Development	
Similarities	Unique ideas
• Like sustainability, experts described equity was a major component of sustainable development	• Experts A and B viewed sustainability as improvement of quality over time • Experts C and D viewed sustainable development as a subset of sustainability

Table 4.1 continued here

Students' Knowledge of Sustainability	
Similarities	Unique ideas
• All the experts noted that students had limited knowledge of sustainability in general • All experts noted that students' knowledge of sustainability was based heavily on the environmental aspect • All experts noted that interest was an important factor in determining students' engagement in the sustainability class. • All experts noted that students who are genuinely actively sought to enroll in sustainability classes	• Expert B described gender differences in students' engagement sustainability classes • Expert D noted that graduate students were more willing to engage in long term sustainability projects than undergraduate students

Measuring Students' Knowledge of Sustainability	
Similarities	Unique ideas
• All the experts tended to measure students' knowledge about sustainability • All the experts expressed that strong emphasis is needed on students' ability to think critically about sustainability • All the experts explained that critical thinking about sustainability is difficult to evaluate	

Therefore, to answer R1 "How do sustainability experts conceptualize sustainability?" I came to the conclusion that based on the findings from qualitative interviews experts conceptualize sustainability as a quality of societies that involves the judicious and equitable use of the earth's resources so current and future generations are able to survive and thrive economically and socially without brining harm to the earth's ability to sustain life.

However, while they agreed on these points, the experts varied in the extent to which they emphasized the different aspects of sustainability (ecology, economy, and social). The experts also varied in their views about sustainability vs sustainable development, except for the idea that equity is crucial for encouraging sustainability development. The experts all shared the sentiment that students had a rudimentary conceptualization of sustainability rooted on environmental issues. The experts tended to measure students' knowledge of sustainability as opposed to their ability to think critically about sustainability mainly due to the difficulty of measuring students' ability to think critically about sustainability. In their attempts to measure students' ability to think critically about sustainability, the experts utilized strategies such projects and peer-reviews to engage students in critical thinking about sustainability.

Given these considerations, utilizing experts' conceptualization of sustainability and the literature review, for the purpose of developing the scale in this study, sustainability means: the earth's resources are utilized in an equitable and judicious manner; current and future generations are able to survive and thrive economically and socially; no harm is brought to the earth's ability to sustain life. Applying this definition of sustainability, the Critical Thinking about Sustainability Scale (CTSS) was designed to measure students' ability to think critically about sustainability.

4.2 Evaluation of Construct Validity

This section provides the results of the evaluation of the extent to which the CTSS measures students' ability to think critically about the three aspects of sustainability (RQ2). The section begins with a description of the scoring reliability and score distribution, followed by the results of the EFA, the 2PL GRM and ends with some recommendations for revising the CTSS.

4.2.1 Scoring Reliability and Score Distribution

Tables 4.2, 4.3, and 4.4 depict the intra-rater reliability of the scoring of student responses, the distribution of scores on the scenarios and the summary statistics for each scenario. To assess the intra-rater reliability of the scoring, I computed Cohen's Kappa (κ) for each of the three scenarios using the criteria: < 0 less than chance agreement, 0.01 – 0.20 slight agreement, 0.21 – 0.40 fair agreement, 0.40 – 0.60 moderate agreement, 0.61 – 0.80 substantial agreement, and 0.81 – 0.99 almost perfect agreement (Viera & Garrett, 2005). I computed the statistics by randomly sampling 10% of the 151 students (or 15 student responses) and rescoring their responses to their questions a second time approximately 1 month after the initial scoring and compared the score responses from the random sample and the original scores from the data. Table 4.2 provides the intra-rater reliability estimates for the three scenarios.

Table 4.2
Cohen's Kappa for the three Scenarios in the CTSS

	Cohen's Kappa		
Question	Keeping Foods and the Climate Cool	Broken Bulb Dreams	Ode to Ogallala
Question 1	.76	.68	.76
Question 2	1.00	1.00	1.00
Question 3	.80	.79	.77
Question 4	.74	.78	.70
Question 5	.71	.66	.76
Question 6	.61	.29	.42

For Question 1 of the KFCC scenario there was substantial agreement in the scoring from the first to second sample. Question 2 had the highest Kappa statistic of the six questions scored probably because it only required students to provide additional pieces of information required to consider the sustainability issues. Therefore, the scoring procedure was a simple matter of counting the number of pieces of information provided by the students. In contrast, Question 6 had the lowest Kappa statistic for all scenarios and had the least consistent scoring results. This result was most likely due to a lack of clarity about what constituted a detailed response to the question and thus resulted in variations in the scoring across the two time points. All other questions showed moderate to substantial agreement in the rating scores, indicating consistency of the scoring across the time points. The pattern of the scoring agreement for the OTO scenario was similar to that of the KFCC scenario except for Question 6 where there was moderate agreement (κ = 0.42) in the scoring as opposed to fair agreement (κ = 0.29) on the KFCC scenario. The BBD scenario had a similar pattern of scoring as the other two scenarios, except that for Question 6 there was substantial scoring agreement between the samples (κ = 0.61), as opposed moderate agreement for the OTO scenario ($\kappa = 0.42$), and fair agreement for the KFCC scenario ($\kappa = 0.29$).

Table 4.3 reports frequency distributions for individual Questions to each Scenario. The evaluation criteria for the questions are: if 50% or more of the students obtaining a score of 1, 2 or 3 points; and within an optimal range of 40% to 60%. If less than 10% of the students obtain a score of 1 point then the question is very easy. If between 10% - 39% of students obtained a score of 1 point then the question is easy. If between 61% – 90% of students obtain a score of 1 then the question is considered difficult and if greater than 90% of students obtain a score of 1 then the question is very difficult. Questions with less than 10% of the students obtaining a score of 3 points are very difficult. Questions with between 10% - 39% of students obtaining a score of 3 points are difficult. Questions with between 61% - 90% of students obtaining a score of 3 points is considered easy and if greater than 90% of students obtain a score of 3 points then the question is considered very easy.

Table 4.3
Frequency Distributions for Individual Questions to each Scenario (*N* = 151)

Keeping Foods and the Climate Clean						
Score	1		2		3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Question 1	126	83.4%	21	13.9%	4	2.6%
Question 2	87	57.6%	32	21.2%	32	21.2%
Question 3	63	41.7%	69	54.7%	19	12.6%
Question 4	42	27.8%	35	23.2%	74	49.0%
Question 5	60	39.7%	76	50.3%	15	9.9%
Question 6	94	62.3%	30	19.9%	27	17.9%
Broken Bulb Dreams						
Score	1		2		3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Question 1	108	71.5%	39	25.8%	4	2.6%
Question 2	87	57.6%	37	24.5%	27	17.9%
Question 3	79	52.3%	60	39.7%	12	7.9%
Question 4	53	35.1%	48	31.8%	50	33.1%
Question 5	77	51.0%	69	45.7%	5	3.3%
Question 6	126	83.4%	22	14.6%	3	2.0%
Ode to Ogallala						
Score	1		2		3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Question 1	108	128	84.8%	21	13.9%	2
Question 2	87	79	52.3%	42	27.8%	30
Question 3	79	82	54.3%	37	24.5%	32
Question 4	53	59	39.1%	37	24.5%	55
Question 5	77	76	50.3%	55	36.4%	20
Question 6	126	92	60.9%	24	15.9%	35

Note. Question 1 evaluated students' analytical skills; Question 2 evaluated students' skill in identifying and applying information; Question 3 evaluated students' problem solving skills; Question 4 evaluated students' identification, analysis and evaluation skills; Question 5 looked at students' skills of evaluation; Question 6 evaluated students' metacognitive skills.

On the KFCC scenario, students scored lowest on Question 1 (analysis skills) and Question 6 (metacognitive skills). These two questions had the highest proportion of students obtaining a score of 1 point (i.e., 83.4% and 62.3%, respectively) indicating that these questions were difficult for students. Questions 3 (problem solving skills) and Question 5 (evaluation skills) had the greatest proportion of students obtaining a score of 2 (i.e., 54.7% and 50.3%, respectively) and the lowest proportion of students obtaining a score of 3 points (i.e., 12.6% and 9.9%, respectively) indicating that these two items were at the optimal difficulty. Questions 2 (identification skills) had 57.6% of students scoring 1 point with 21.% scoring 3 points indicating that the question was difficult. Question 4 (identification and analysis skills) 27.8% of students obtaining a score of 1 point and 49.0% of students obtaining a score of 3 points indicating that Question 4 was a relatively easy item.

The OTO scenario had a similar pattern of responses to the KFCC scenario. Questions 1 and 6 had the greatest proportion of students obtaining a score of 1 point (i.e., 84.8% and 60.9%, respectively) indicating that they were difficult items. While Questions 3 and 5 had a greater proportion of students obtaining a score of 2 points 24.5% and 36.4%, respectively. These questions had a relatively low proportion of students obtaining a score of 3 points (i.e., 21.2% and 13.2%, respectively) indicating that these items were relatively difficult for students. Questions 2 and 4 had a high proportion of students obtaining a score of 3 points (i.e., 19.9% and 36.4%, respectively) indicating that these questions were relatively easy.

Similarly, for the BBD scenario Questions 1 and 6 had the highest proportion of students obtaining a score of 1 (i.e., 71.5% and 83.4%, respectively) coupled with the lowest proportion of students obtaining a score of 3 points (i.e., 2.6% and 2.0%, respectively) indicating that these questions were very difficult for students. Of the three scenarios, Questions 3 and 5 had the lowest proportion of students obtaining a score of 3 points (i.e., 7.9% and 3.3%, respectively) and the greatest proportion of students obtaining a score of 2 points (i.e., 39.7% and 45.7%, respectively) indicating that these items difficult for students. Questions 2 and 4 relatively easy items

In Table 4.4 the means (*M*), standard deviations (*SD*), minimum (min) and maximum (max) values, skewness and kurtosis values were examined for plausibility.

Table 4.4
Summary Statistics for the Three Scenarios

	Mean	SD	Min	Max	Skewness	Kurtosis
KFCC Scenario	1.67	0.45	1.00	2.83	0.26	-0.76
BBD Scenario	1.53	0.40	1.00	2.83	0.82	0.59
OTO Scenario	1.62	0.48	1.00	2.83	0.48	-0.61

The mean score on the KFCC scenario was 1.67 points with a standard deviation of 0.45, which were reasonable given distribution of scores. The min and max scores (i.e., 1 and 2.83 points, respectively) were not surprising since several students obtained a score of 1 point on all the questions and no student was able to get a perfect score on KFCC scenario. The skewness and kurtosis values were both well within the ranges of acceptability (i.e., 3 and 10, respectively) as suggested by Kline (2011). A very similar pattern of responses were obtained for the OTO scenario which had marginally lower mean score ($M = 1.62$) and marginally higher standard deviation ($SD = 0.48$). The min and max scores were identical to the KFCC scenario and the skewness and kurtosis values were both within acceptable ranges. Of the three scenarios the BBD scenario had the lowest mean score ($M = 1.53$) and standard deviation ($SD = 0.40$), while having identical min and max scores to the other scenarios and acceptable skewness and kurtosis values.

4.3 Evaluation of CTSS Dimensionality with Exploratory Factor Analysis

Following the evaluation of the scoring reliability and distribution of scores across the scenarios, I evaluated the reliability of the questions in each scenario followed by an exploratory factor analysis (EFA) on the data for each scenario separately. Table 4.5 shows the results of the reliability statistics for the three scenarios. For the KFCC scenario, Cronbach's Alpha (α) was less than adequate (Nunnally, 1978). The item-total correlation (ITC) for Question 1 was 0.19 indicating the need for substantial revision or deletion that would result in an increase in α to 0.71. Question 2 showed very good ITC (0.40) and would cause a small reduction in α to 0.67 is deleted.

Table 4.5

Reliability Estimates for the Three Scenarios

	KFCC		BBD		OTO	
	Item-Total Correlation	Alpha if Item Deleted	Item-Total Correlation	Alpha if Item Deleted	Item-Total Correlation	Alpha if Item Deleted
Question 1	0.19	0.71	0.35	0.65	0.21	0.72
Question 2	0.40	0.67	0.42	0.64	0.41	0.67
Question 3	0.55	0.61	0.50	0.60	0.51	0.64
Question 4	0.42	0.66	0.31	0.69	0.46	0.66
Question 5	0.56	0.61	0.50	0.61	0.52	0.64
Question 6	0.44	0.65	0.48	0.63	0.51	0.63
Scenario Alpha	0.69		0.70		0.70	

Question 3 also had a very good ITC (0.55) and its removal from the scenario would cause a relatively large drop in α to 0.61. Question 4 had a similar ITC value (0.42) to Question 2 (0.40) and if deled from the scenario would cause a small drop in α (0.66). Question 5 had a ITC value (0.56) similar to Question 3 (0.55) and its removal from the scenario would cause a similarly large drop in α to 0.61. Question 6 had a similar ITC value to Questions 2 (0.40) and 4 (0.42) and its removal from the scenario would result in a relatively moderate drop in α to 0.65.

The results were a little different for the BBD scenario. Unlike the KFCC scenario, the BBD scenario had adequate reliability. Question 1 had a reasonable ITC statistic (0.35) and if deleted from the scenario would result in a drop in α to 0.65. Question 4 on the BBD scenario had reasonable ITC (0.31) as opposed to very good (0.42) on the KFCC scenario and if deleted from the scale would cause a slight drop in α to 0.69. All other ITC results were consistent with the KFCC scenario. The results of OTO scenario were consistent with those of the KFCC scenario except for Question 1 that had marginal ITC (0.21) instead of the poor ITC of Question 1 in the KFCC scenario (0.19)

Following the evaluation of the reliability statistics the polychoric correlations among the questions in the three scenarios were evaluated and depicted in Tables 4.6, 4.7, and 4.8. The correlations for the Keeping Foods and the Climate Cool (KFCC) scenario ranged from 0.16 – 0.64; for the Ode to Ogallala (OTO) scenario correlations ranged from 0.23 – 0.65; for the BBD scenario the correlations ranged from 0.16 – 0.61. These results indicated a small to moderate correlations among the questions. There were significant correlations among Questions 3,4,5 and 6 ($p < 0.05$) which was expected as these questions measured students' problem solving skills (Question 1), skills in identifying, analyzing, and evaluating information (Question 4), their evaluative skills (Question 5) and their metacognitive skills (Question 6) all of which are higher order thinking skills associated with sustainability.

There was a significant correlation between Questions 1 and 6 which was expected since Question 1 assessed students' skills of analysis and Question 6 assessed their metacognitive skills. There were also significant correlations among items 2,3,4,5 and 6 of the KFCC scenario indicating a relationship between identifying information (Question 2) and higher order thinking skills (Questions 3, 4, 5 and 6). The pattern of correlations on the BBD scenario was similar to the KFCC scenario except in two ways. First, in the BBD scenario Question 2 did not significantly correlate with Question 4 and 2) Question 1 was significantly correlated to all other items except Question 4, which was not the case for the KFCC scenario. The pattern of correlations for the OTO differed from the BBD scenario only to the extent that Question 2 was not significantly correlated with Question 5 unlike the BBD scenario and Question 1 was uncorrelated with either Questions 2 or 6. Overall, the patterns of correlations indicated the presence of a single factor.

Table 4.6
Polychoric Correlation for Keeping Foods and the Climate Cool Scenario

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
Question 1	1					
Question 2	.16	1				
Question 3	.16	.46*	1			
Question 4	.22	.37*	.45*	1		
Question 5	.22	.43*	.64*	.41*	1	
Question 6	.31*	.30*	.48*	.40*	.53*	1

Note. * indicates $p < .05$

Table 4.7
Polychoric Correlation for Broken Bulb Dreams Scenario

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
Question 1	1					
Question 2	.47*	1				
Question 3	.31*	.49*	1			
Question 4	.23	.16	.31*	1		
Question 5	.31*	.31*	.57*	.46*	1	
Question 6	.37*	.56*	.55*	.37*	.61*	1

Note. * indicates $p < 0.05$

Table 4.8
Polychoric Correlation for Ode to Ogallala Scenario

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6
Question 1	1					
Question 2	.23	1				
Question 3	.37*	.41*	1			
Question 4	.36*	.35*	.37*	1		
Question 5	.30*	.26	.65*	.51*	1	
Question 6	.25	.53*	.45*	.50*	.48*	1

Note. * indicates $p < 0.05$

Following the examination of the correlation structure, I evaluated the scree plot and the eigenvalues of the data. Figures 4.1, 4.2 and 4.3 show the scree plot of the eigenvalues where the x-axis contains the number of factors while the y-axis contains the eigenvalue estimates. For example, in Figure 4.1, the scree plot for the KFCC scenario, the discontinuity in the eigenvalues occurred at the second factor and therefore there is only one factor prior the discontinuity indicating that a one-factor solution is plausible. Moreover, the single factor for the KFCC scenario had an eigenvalue of 2.92, well above the cut-off criteria of 1. The scree plots in Figures 4.2 for the BBD scenario and 4.3 for OTO scenario displayed similar pattern found in Figure 4.1. Similarly, the eigenvalue estimates for the BBD (3.07) and OTO (3.04) scenarios were well above the cut-off value of 1. Thus, Figures 4.1, 4.2 and 4.3 indicate a one-factor solution and that factor accounts for a substantial amount of variation in the questions of the CTSS.

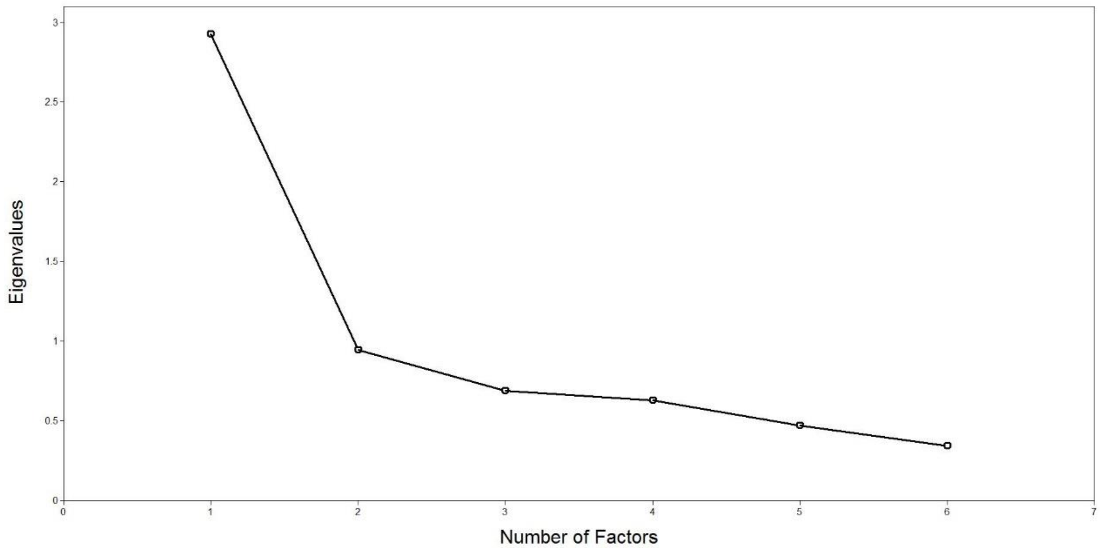


Figure 4.1
Scree plot for the KFCC Scenario

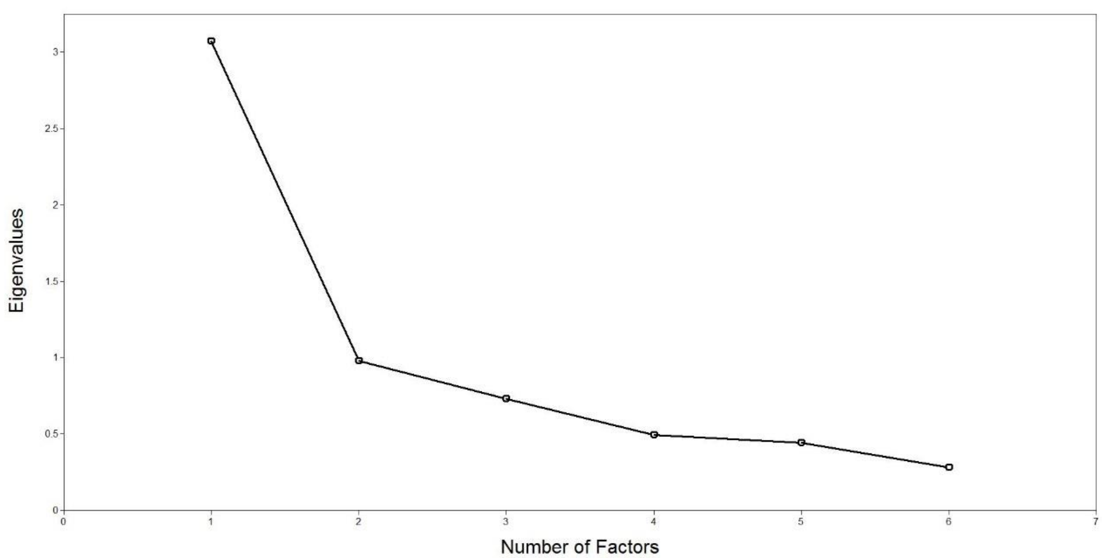


Figure 4.2

Scree plot for BBD Scenario

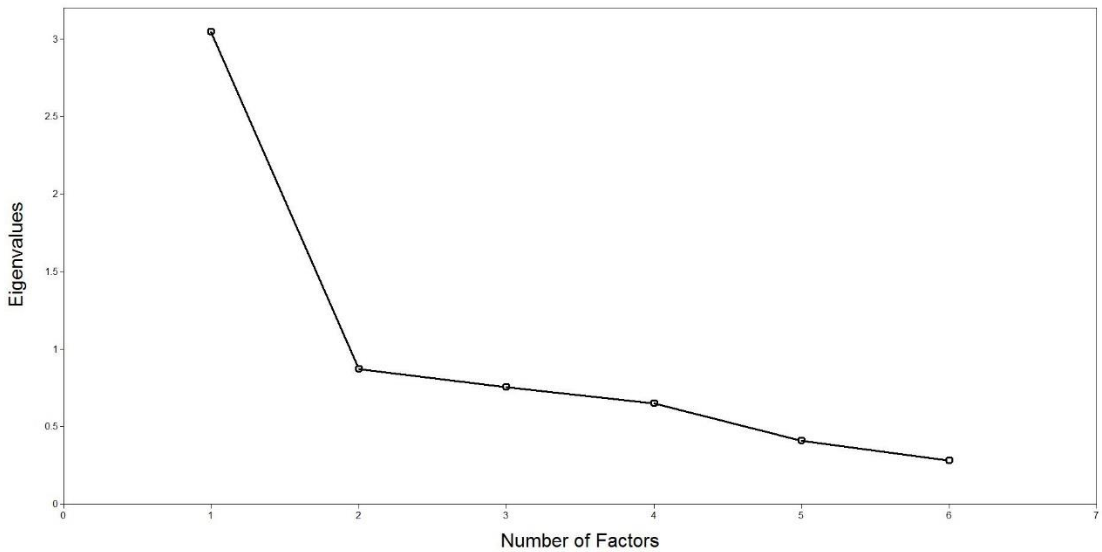


Figure 4.3
Scree plot for the OTO Scenario

Based on the results of the correlation matrices, scree plots, and eigenvalues I fitted a one-factor EFA model to the data. Table 4.9 depicts the model fit statistics derived from the EFA. The fit statistics indicated that the hypothesized one factor model provided an excellent to the observed data for the three scenarios. For example for the results indicate that EFA model for the KFCC scenario fit the data well χ^2 (9, N = 151) = 4.85, p = 0.85; RMSEA = 0.00 90% C.I. [0.00, 0.05], CFI = 1.00, TLI = 1.00, and SRMR = 0.04. Like the KFCC scenario, the model fit results for the BBD scenario also showed excellent fit of the model to the data χ^2 (9, N = 151) = 15.62, p = 0.08; RMSEA = 0.07, 90% C.I. [0.00, 0.13], CFI = 0.97, TLI = 0.94), and SRMR = 0.07. Similarly, the model fit results for the OTO scenario showed excellent fit to the data as the KFCC and BBD scenarios χ^2 (9, N = 151) = 16.35, p = 0.06; RMSEA = 0.07, 90% C.I. [0.00, 0.13], CFI =

0.97, TLI = 0.95, and SRMR = 0.07. Thus, I conclude that one-factor model is appropriate for analyzing data on CTSS.

Table 4.9
Model Statistics for the Exploratory Factor Analysis of the CTSS

Scenario	χ^2	*P*	RMSEA	90% CI RMSEA	CFI	TLI	SRMR
KFCC	4.85	0.85	0.00	0.00, 0.05	1.00	1.00	0.04
BBD	15.62	0.08	0.07	0.00, 0.13	0.97	0.94	0.07
OTO	16.35	0.06	0.07	0.00, 0.13	0.97	0.95	0.07

Note. χ^2 measured at $df = 9$; $N = 151$

Table 4.10 shows the results of the EFA model where the columns indicate the factor loadings for each item on the respective scenario. The factor loadings represent the relationship between each question and the underlying latent construct (Brown, 2006). Therefore, the higher the factor loading the stronger the relationship between the question and the latent construct. The variance on question column shows the proportion of the variance in the question explained by the factor. The residual variance is equal to 1 minus the proportion of variance in the question explained by the factor and denotes the remaining variance after account for the presence of the factor. The total variance explained row, describes the proportion of the variance in all variables attributable to the latent construct.

Table 4.10
Results of the EFA for the Three Scenarios

Scenario	Factor Loading	Communality	Residual variance	Explained Variance
		KFCC		
Question 1	0.30*	0.09	0.91	40%
Question 2	0.56*	0.31	0.69	
Question 3	0.79*	0.62	0.38	
Question 4	0.58*	0.33	0.67	
Question 5	0.80*	0.64	0.36	
Question 6	0.64*	0.41	0.59	
		BBD		
	Factor Loading	Communality	Residual variance	Explained Variance
Question 1	0.51*	0.26	0.74	38%
Question 2	0.62*	0.39	0.61	
Question 3	0.72*	0.52	0.48	
Question 4	0.47*	0.22	0.78	
Question 5	0.73*	0.53	0.47	
Question 6	0.82*	0.67	0.33	
		OTO		
	Factor Loading	Communality	Residual variance	Explained Variance
Question 1	0.45*	0.20	0.80	43%
Question 2	0.55*	0.31	0.69	
Question 3	0.75*	0.57	0.43	
Question 4	0.63*	0.40	0.60	
Question 5	0.77*	0.59	0.41	
Question 6	0.71*	0.51	0.49	

Note. Communality is the amount of variance explained in the question by the factor. * indicates statistically significant factor loading , p <.05

As Table 4.10 shows for the KFCC scenario, the factor loadings ranged from 0.30 – 0.80 and were all statistically significant ($p < 0.05$). Question 1 had the lowest factor loading, which was below the cut-off criteria of 0.32. Unsurprisingly, the factor explained the least amount of variance on this question (9%) and the question had the highest residual variance (91%). Question 1 examined students' ability to discern the three aspects of sustainability given the information in the text and had the greatest proportion of students who obtained a score of 1 (83.4%)—the lowest score possible and lowest proportion of students who obtained a score of 3 (2.6%) indicating the question was difficult. From Table 4.5 it is evident that this question had the lowest ITC (0.19) indicating that it was a poor question as well. Question 6 was another question to take note of because it had the second highest proportion of students obtaining a score of 1 (62.3%) making it a fairly difficult question. Question 6 had a moderate factor loading (0.64) and aimed to evaluate students' metacognitive skills.

Questions 3 and 5 had the highest factor loadings 0.79 and 0.80, respectively. Unsurprisingly, the factor loaded strongly onto these questions since they require students to display strong analysis, evaluation, and synthesis skills to answer these questions. These questions had the highest ITC of the six questions (i.e., 0.55 and 0.56, respectively) indicating that they were good questions. Question 3 aimed to measure students' problem solving ability given limited information and Question 5 examined students' evaluative skills. Additionally, Questions 3 and 5 had the greatest proportion of students obtaining a score 2 (i.e., 54.7% and 50.3%, respectively) within the optimal range for student responses.

Questions 2 (identifying information) and 4 (identifying, analyzing, and evaluating information) had similar factor loadings (i.e., 0.56 and 0.58, respectively) and the factor explained a relatively low proportion of variance on the items 0.31 and 0.33 respectively. These questions also had the greatest proportion of students obtaining a score of 3 (i.e., 21.2% and 49%, respectively) indicating that they were easier than the other questions. These two questions also had very good ITC 0.40 and 0.42, respectively. Apart from Question 1, these two questions had the lowest factor loadings of the six questions. However, despite their very good ITC and strong factor loadings these questions were the least cognitively demanding questions of the test. The underlying construct explained 40% of the variation in students' responses to the questions on the KFCC scenario.

A similar pattern of responses was found for the BBD scenario with factor loadings ranging from 0.47 - 0.82. Question 1 for the BBD scenario had the second highest proportion of students obtaining a score of 1 point, 71.5% indicating the item was difficult. The factor explained the least amount of variance on Question 1 (0.26) and had a reasonable ITC of 0.31 indicating that the question could be improved with revisions. Question 6 had the greatest proportion of students obtaining a score of 1 (83.4%) indicating that the item was difficult and had very good ITC 0.48 (Table 4.5).

Like the KFCC scenario, Questions 3 and 5 of the BBD scenarios similarly high factor loadings (i.e., 0.72 and 0.73, respectively) and the factor explained the second and third highest proportion of variance on the items 0.52 and 0.53, respectively. Both questions had the highest proportion of students obtaining a score of 2 points (i.e., 39.7% and 45.7%, respectively) and were within the optimal range for the scenario. Questions 3 and 5 also had very good discrimination values 0.50 and 0.50, respectively. Questions 2 and 4 had moderate factor loadings (i.e., 0.62 and 0.47, respectively) and like Questions 2 and 4 of the KFCC scenario, they had the highest proportion of students obtaining a score of 3 points 17.9% and 33.1%, indicating that these items were relatively easy for students. The ITC for Question 4 was 0.31 indicating that it can be improved with revisions. The factor accounted for 38% of the variations in the items.

The results of the OTO scenario were similar to that of the KFCC scenario in that Question 1 had lowest factor loading of the questions in the scenario with the factor explaining the least amount of variance on the question. Question 1 had the greatest proportion of students obtaining a score of 1 (84.8%) on the scenario and a very low proportion of students obtaining a score 3 (1.3%) indicating that the item was difficult for students. The results for Question 6 show that the factor loaded strongly on the item (0.71) and had a high proportion of students obtaining a score of 1 (60.9%) and similar finding to the BBD and KFCC scenarios. Question 6 also had the largest proportion of students obtaining a score of 3 points 23.2 % indicating it was easier on the OTO scenario as opposed to the BBD or KFCC scenarios.

Like the BBD and KFCC scenarios, Questions 3 and 5 of the OTO both had strong factor loadings 0.75 and 0.77, respectively. These questions also had a larger proportion of students obtaining a score of 3, 21.2% and 13.2% when compared to the other scenarios indicating that these questions were easier for students to answer on the OTO scenario when compared to the KFCC and BBD scenarios. Questions 2 and 4 of the OTO scenario showed similar patterns of results as the other scenarios. Both questions had moderate loadings (i.e., 0.55 and 0.63, respectively) and had a high proportion of students obtaining a score of 3 points 19.9% and 36.4%, indicating that they were relatively easy items. Of all the scenarios, the factor accounted for most of the variation in the OTO scenario (43%).

Overall, across the three scenarios, several clear patterns emerged. All six questions in each of the scenarios loaded significantly ($p < .05$) onto the factor, defined as critical thinking ability about sustainability, with moderate to high loadings. As one noticed from the table the factor loadings across the scenarios were not equal indicating that the students’ responses to the questions on the scenario may have been affected by the content within each scenario. Across all scenarios, Questions 1 and 6 are the most difficult given the large proportions of students obtaining a score of 1 point. Question 1 in particular seemed to be the least suitable question given its relatively low ITC (Table 4.5) and factor loadings (Table 4.10) and the corresponding increase in the reliability of the scale should it be removed. Another pattern that emerged was that the easier questions that measured students’ lower level thinking skills (Questions 2 and 4); tended to have lower ITC (Table 4.5), lower factor loadings (Table 4.10); and a greater proportion of students obtaining a score 3 points (Table 4.4) compared to Questions 3, 5 and 6 that assessed students’ higher order thinking skills. Moreover, it is noteworthy that even Question 6 that was most affected by the scoring (lowest Kappa values Table 4.3) had greater ITC and factor loading values than Questions 2 and 4. This underscores the need to either substantially revise or remove Questions 2 and 4 from the scale.

4.4 Evaluation of CTSS item quality with item response theory

The EFA analysis in the preceding section established the question within the scenarios measured one construct. Therefore, I used 2PL GRM to the data. The results of this analysis are provided in this section beginning with an evaluation of the thresholds and slopes, followed by an inspection of the IIFs and TIFs and ending with the model fit statistics and summary of recommendations for editing and potentially changing questions.

4.4.1 Thresholds and slopes

Table 4.11 depicts the model fit statistics derived from the 2PL graded response model. To determine the adequacy of the fit of the model to the data the Chi-square statistic ($\chi2$) was evaluated where a non-significant $\chi2$ indicates adequate fit and significant $\chi2$ indicates less than adequate fit between the observed and expected frequencies of the model. The fit statistics indicated non-significant $\chi2$ which means that the 2PL GRM fit the data adequately.

Table 4.11
Model Fit Statistics for the 2PL Graded-Response Model for the CTSS

Scenario	χ^2	*df*	*p*
Keeping Foods and the Climate Clean	516.20	708	0.99
Broken Bulb Dreams	329.43	708	0.99
Ode to Ogallala	457.43	468	0.63

Note. * indicates statistically significant χ^2

Table 4.12 depicts the results of the 2PL graded-response model for the three scenarios. As mentioned earlier, the slope parameter represents the question's ability to distinguish between examinees with different latent ability. The threshold parameters represent the ability level necessary to respond above threshold with 0.50 probability. The first threshold (T1) indicates the latent ability that determines whether a student will get a score of 1 or 2 points. Therefore, this means that a student with a latent ability above T1 will obtain a score of 2 points and a student with latent ability lower than T1 will obtain a score of 1 point on CTSS (Adams, Wu, & Wilson, 2012). Threshold 2 (T2) indicates the latent ability that intersects the point where a student would get either a score of 2 or 3 points. This means that a student with latent ability greater than T2 will obtain a score of 3 points (Adams et al., 2012). Under ideal circumstances, a researcher would want an equal slopes and equidistant thresholds for all questions. Equal slopes and equidistant thresholds would indicate that the questions discriminates equally among students between adjacent categories, therefore any observed differences in students' responses would be due to differences in the levels of the latent trait of the students (Baker, 2001).

In the KFCC scenario one can see that for Question 1 T1, a student with an ability level greater than 1.77 would obtain a score of 2 points and a student with ability lower than 1.77 will obtain a score of 1 point. Similarly, a student with ability above 4.05 will obtain a score of 3 points indicating it is very difficult to get a score of 3. However, the slope estimate for Question 1 was low indicating that the response categories were not good at differentiating along the range of the latent ability. This is consistent with the fact that a significantly greater number of students obtained a score of 1 (83.4%) as opposed to two points (13.9%) or 3 points (2.6%). In other words more 5 times as many students obtained a score of 1 as opposed to a score of 2 and as much as 5 times as many students obtained a score of 2 (13.9%) as opposed to a score of 3 (2.6%). It is also worth noting that Question 1 had the lowest factor loading (0.30) on Table 4.10 and the lowest discrimination parameter (0.19) on Table 4.5 of all the questions in the scenario making it necessary to revise Question 1 (a measure of students analysis skills) for further use in the scale.

Table 4.12
Threshold and Slope Estimates for the 2PL Graded-Response IRT Model for the CTSS

Scenario/Questions	T1(*SE*)	T2(*SE*)	Slope
KFCC			
Question 1	1.77 (0.26)	4.05 (0.59)	0.58 (0.32)
Question 2	0.37 (0.21)	1.66 (0.24)	1.12 (0.25)
Question 3	-0.56 (0.31)	3.46 (0.64)	2.37 (0.54)
Question 4	-1.17 (0.23)	0.09 (0.21)	1.14 (0.27)
Question 5	-0.71 (0.32)	3.74 (0.67)	2.34 (0.56)
Question 6	0.73 (0.25)	2.10 (0.33)	1.47 (0.32)
BBD			
Question 1	1.10 (0.23)	4.07 (0.56)	1.01 (0.33)
Question 2	0.37 (0.22)	1.92 (0.29)	1.26 (0.34)
Question 3	0.13 (0.25)	3.53 (0.53)	1.77 (0.43)
Question 4	-0.73 (0.20)	0.83 (0.21)	0.93 (0.25)
Question 5	0.05 (0.26)	4.88 (0.77)	1.90 (0.45)
Question 6	2.85 (0.64)	6.32 (1.22)	2.40 (0.69)
OTO			
Question 1	3.06 (0.50)	-	1.01 (0.49)
Question 2	0.14 (0.20)	1.66 (0.27)	1.10 (0.30)
Question 3	0.22 (0.27)	2.28 (0.44)	2.00 (0.54)
Question 4	-0.69 (0.23)	0.83 (0.25)	1.42 (0.35)
Question 5	0.02 (0.28)	3.37 (0.63)	2.19 (0.61)
Question 6	0.80 (0.28)	2.04 (0.38)	1.68 (0.46)

Note. Question 1 of the Oto Scenario only had one threshold because T2 is too high and could not be estimated given the data ; *SE*s in parentheses.

Another question that flagged through previous analyses was Question 6. The result of IRT analysis indicated that the thresholds of Question 6 had relatively high T1 (0.73) and T2 (2.10), respectively, suggesting that students should possess very high critical thinking skills to obtain a score of 3 points. This means that Question 6 discriminated among students on the lower end of the latent ability distribution because students' responses to the adjacent difficulty categories 1 and 2 are further apart than categories 2 and 3. This is quite evident from Table 4.4 where almost equal proportions of students obtained a score of 2 (19.9%) or 3 points (17.9%) while more than 3 times as many students obtained a score of 1 point (62.3%) as opposed to 2 points (17.9%). On Table 4.10 one can see that Question 6 had the third highest factor loading (0.64) and on Table 4.5 Question 6 had the third highest item-total correlation parameter 0.44. This indicates that Question 6 can be a potentially good item for the scenario.

The findings for Questions 3 (problem solving) and 5 (evaluation) were quite different from Questions 1 and 6 above. Both questions had high T2 values but low T1 values indicating that it was much easier to obtain a score of 2 points than 3 points. Consequently, the distance for a student to move from a score of 1 to 2 points is not equal to the distance required for the student to move from 2 to 3 points in terms of their critical thinking ability. This means that T1 and T2 would be more effective in discriminating students higher on the latent ability in the sample, because their responses to the adjacent higher difficulty categories (2 and 3) could be very different. When one looks at Table 4.4 it is easy to see that almost equal proportions of students obtained as score of 1 or 2 points—hence the inability for Questions 3 and 5 to differentiate among the two categories and thus the low T1.

However, when we look at categories 2 and 3 one can see that more than 3 times as many students obtained a score of 2 (54.7%) as opposed to a score of 3 (12.6%) on Question 3. On Question 5 more than 5 times as many students obtained a score of 2 (50.3%) as opposed to a score of 3 (9.9%). Hence, the ability of these Questions to differentiate among students higher on the latent ability distribution and the corresponding high slope parameters for the two questions 2.37(Question 3) and 2.34(Question 5). It also noteworthy that these two questions had the highest factor loading Question 3 (0.79) and Question 5 (0.80) in Table 4.10 and the highest ITC

Question 3 (0.55) and Question 5 (0.56) in Table 4.5 of the scenario. The above results indicate that Questions 3 and 5 are suitable for the scale.

The results for Questions 2 (skills in identifying information) and 4 (identifying, evaluating and analyzing information) were in contrast to those of Questions 3 and 5. Question 2 had a moderately high T1 and a high T2 values indicating that it was easier to obtain a score of 2 points than 3 points. This means that T1 and T2 would be more effective in discriminating students lower on the latent ability in the sample, because their responses to the adjacent higher difficulty categories (2 and 3) are very different. When one looks at Table 4.4 it is noticeable that for Question 2 more than 2 times as many students obtain a score of 1 (57.6%) than a score of 2 (21.2%); in contrast, an equal proportion of students obtained as score of 2 (21.2%) or 3 (21.2%) points. Hence, the inability for Question 2 to differentiate between categories 2 and 3, thus the relatively low T2 and the moderate slope estimate (1.14). However, when we look at categories 1 and 2 one can see that more than 3 times as many students obtained a score of 1 (57.6%) as opposed to a score of 2 (21.2%) on Question 2. Hence the greater ability of Question 2 to differentiate between categories 1 and 2. Moreover, Table 4.10 shows that Question 2 had a lower factor loading than Questions 3, 5 and 6 and a lower ITC as well. These results show that Questions 2 and 4 are not suitable for the scale.

The pattern of results for the BBD scenario was similar to that of the KFCC scenario except for Question 6 that had high T1 and T2 values. The results of Table 4.4 show that almost 5 times as many students obtained a score of 1(83.4%) than those who scored 2 (14.6%) and more than 7 times as many students obtained a score of 2 (14.6%) as opposed to a score of 3 (2.0%). This accounts for the very high T1 and T2 values and the very high slope value 2.40 that indicates that the question discriminates strongly among students between the adjacent categories. Moreover, the results of Table 4.10 shows that Question 6 had the highest factor loading (0.80) and very good discrimination values (0.48) indicating that the item can be retained in the scale.

The pattern of results for the OTO scenario was similar to that of the KFCC scenario except for Question 1 that had very a high T1 but no T2 indicating that the T2 threshold was so high that the program could not estimate it. The results of Table 4.4 show that more than 5 times as many students obtained a score of 1(84.8%) than those who scored 2 (13.9%), and more than 10 times as many students obtained a score of 2 (13.9%) as opposed to a score of 3 (1.3%). That accounts for the very high T1 (3.06), but like Question 1 of the other two scenarios provides low information despite being able to differentiate between students on the lower categories in that the question. Moreover, the results of Table 4.10 show that Question 1 had the lowest factor loading (0.45) and the lowest ITC of the questions on the OTO scenario.

In summary, across the three scenarios Questions 3 and 5 are best suited to the scale as they measure higher order thinking skills and have good psychometric evidence to support their retention in the scale. Question 6 has the potential to be a very good question pending revisions. However, Question 1 needs major revision as the results show it is not adequate for the scale, but measures an important skill on the scale (students' ability to describe the different aspects of sustainability). Questions 2 and 4 should be extensively revised or replaced since the evidence suggest that they are not suitable for the scale and tended to measure lower level thinking skills.

4.4.2 Item information function (IIF) evaluation

This section presents the results provided by the IIFs used to evaluate the quality of the questions on each scenario depicted in Figures 4.4, 4.5, and 4.6. For the KFCC scenario, it is evident from Figure 4.4 that, consistent with previous evaluations, Question 1 provided the least amount of information of the six questions. This is not surprising since Table 4.12 shows that Question 1 had the lowest slope estimate (0.58), that it had the lowest factor loading (0.30) showed in Table 4.10, and the ITC (0.19) showed in Table 4.5. On the other hand, Questions 3 and 5 provided the most amount of information and all three figures given as indicated by the high bimodal peaks and the very high slope values (i.e., 2.37 and 2.34 respectively) depicted in Table 4.12. Additionally, these questions had the highest factor loadings of the six questions on the scenario (i.e., 0.79 and 0.80 respectively) shown in Table 4.10. They also had the highest ITC 0.55 and 0.56 (Table 4.5). Questions 2, 4 and 6 provided a relatively low amount of information given their low peaks and slope estimates (i.e., 1.12, 1.14, and 1.47, respectively) depicted in Table 4.12. Additionally, Table 4.10 showed that these questions had moderate factor loadings of 0.56, 0.58, and 0.64.

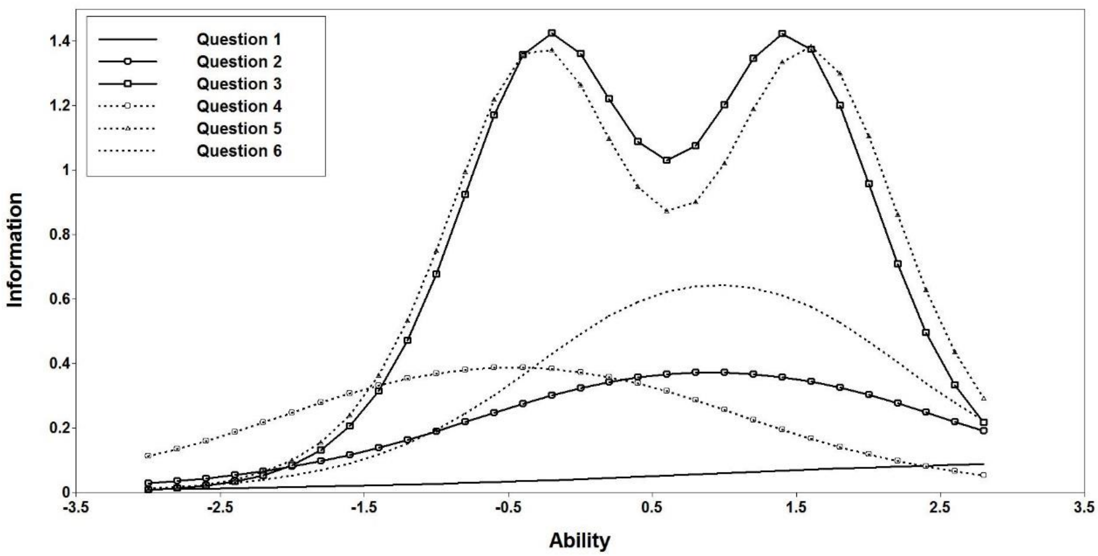


Figure 4.4
Item Information Function for the KFCC Scenario

For the BBD scenario depicted in Figure 4.5 one can see that Questions 1 and 4 provided the least amount of information as denoted by their flat curves and moderate estimates 1.01 and 0.93 depicted in Table 4.12. Moreover, these two questions had the lowest factor loadings of the six questions in the scenario (i.e., 0.51 and 0.47, respectively) shown in Table 4.10, these two question also had the lowest item-total correlation 0.35 and 0.3 (Table 4.5). Question 2 also provided a low amount of information (1.26) as seen in Table 4.12. Question 6 provided the most amount of information and showed bimodal peaks and high slope values 2.40 (Table 4.12). Unexpectedly, Questions 3 and 5 provided the most amount of information (i.e., 1.77 and 1.90, respectively) after Question 6 (Table 4.12). These question also had the highest factor loadings of the six questions 0.72 and 0.73 (Table 4.10) and the highest ITC 0.50 and 0.50 (Table 4.5).

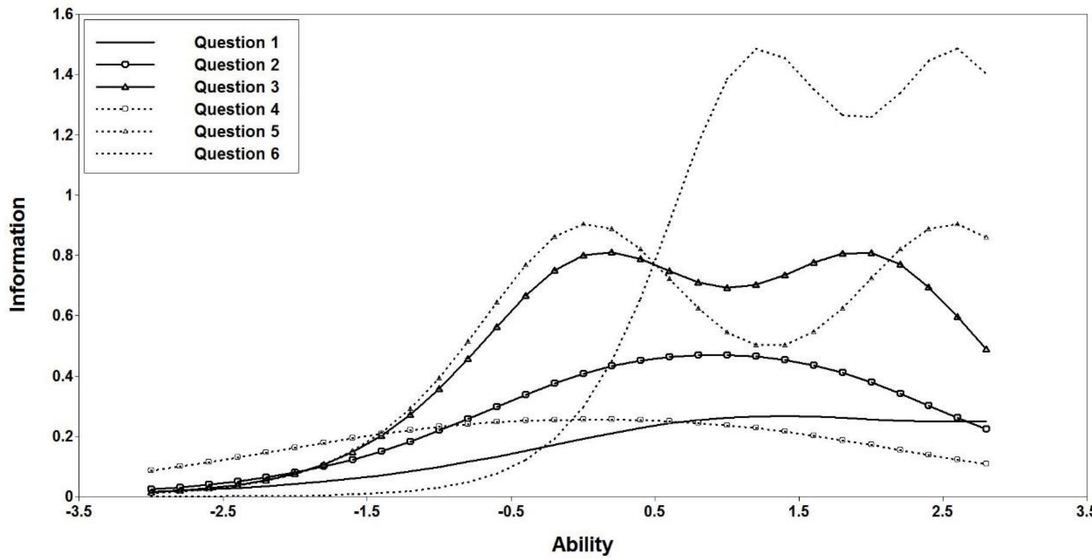


Figure 4.5
Item Information Function for the BBD Scenario

The IIFs for the OTO Figure 4.6 shows a different pattern of functions. As in the other scenarios, Question 1 of the OTO scenario provided the lowest amount of information with a moderate slope of 1.01. Questions 2 and 4 provided more information than Question 1 with slightly higher moderate slopes (i.e., 1.10 and 1.42, respectively) shown in Table 4.12. Unsurprisingly, Questions 3 and 5 provided the most amount of information with very high slopes estimates of 2.00 and 2.19 respectively. The information curve for Question 5 in particular was bimodal indicating that two trait levels provided the maximum information. Question 6 provided the most amount of information after Questions 3 and 5, with a high slope estimate of 1.68.

Overall, from these three graphs one can see a clear trend, Questions 1, 2, and 4 that required limited cognitive effort from students provided the least amount of information. On the other hand, Question 3, 5 and 6 that were more difficult and required higher order thinking skills provided more information (Table 4.12). Additionally, Questions 3, 5, and 6 also had higher ITCs (Table 4.5) and factor loadings (Table 4.10) than Questions 1, 2 and 4. However, it should be noted that despite the fact that Questions 3 and 5 provided the most information, they were distinctly bi-modal which makes it difficult to determine what latent trait level provides the most amount of information.

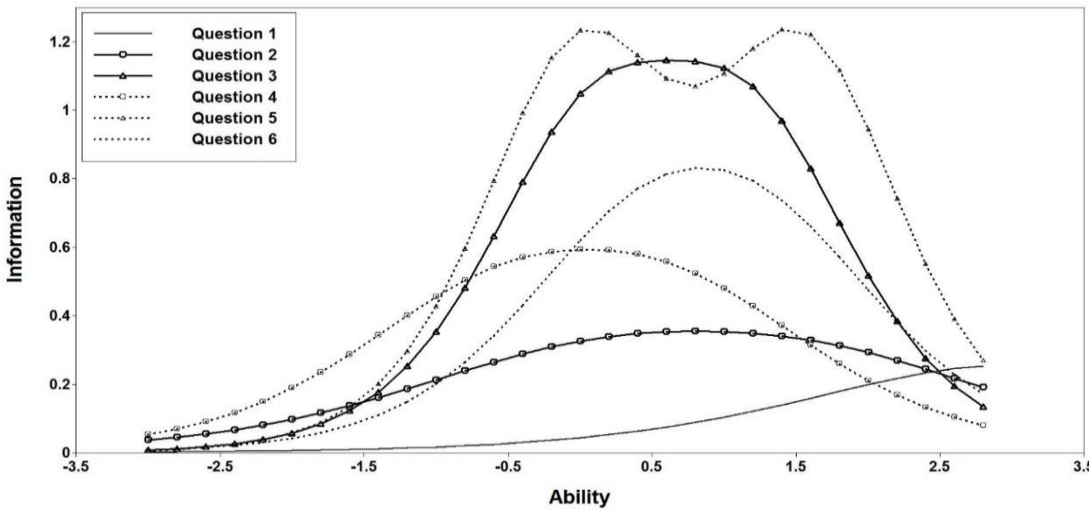


Figure 4.6
Item Information Function for the OTO Scenario

4.4.3 Test information function (TIF) evaluation

Figures 4.7, 4.8 and 4.9 provide the results of the test information functions for the three scenarios. The test information, provided by Figure 4.10 indicates that most of the information on the KFCC scenario centers approximately around ability ranges -0.5 and 1.5 indicating that the KFCC scenario is informative for students of relatively low and high ability levels with a distinctly bimodal shape. The test information, provided by Figure 4.11 indicates that most of the information on the BBD scenario centers approximately on ability range 1.0 and 2.5 and had a flatter function indicting that it discriminates over a wider range of latent trait and bimodal to a lesser extent than the

KFCC scenario. The test information curve, provided by Figure 4.12 indicates that most of the information on the OTO scenario occurs at ability 0 and had a relatively steep curve indicating that it discriminated along a narrower range of latent traits. This stands in contrast to TIF BBD scenario that distinguished among students along a wider latent trait range.

Overall, the BBD scenario provided a little less information at its peak than the KFCC and OTO scenarios while the KFCC and BBD scenarios tended to have bimodal curves particularly the KFCC test information function. The ideal form of the CTSS would provide equal information across the whole range of ability levels on the scale indicated by a horizontal line across the graph. This ideal graph would mean that the scale discriminated among students of various latent traits equally hence, any observed differences in scores would be due to differences in the students' latent ability alone. However, this ideal form of the scale would require including questions with very low discrimination and hence poor precision that would lower the overall quality of the scale (Baker, 2001). However, the desirable shape of the TIF depends on the purpose of using the instrument in practice (de Ayala, 2009; Embretson & Reise, 2000). In this case, the shapes of the TIFs indicate that the CTSS may be best used for students whose latent trait fall on the average to above-average ability levels.

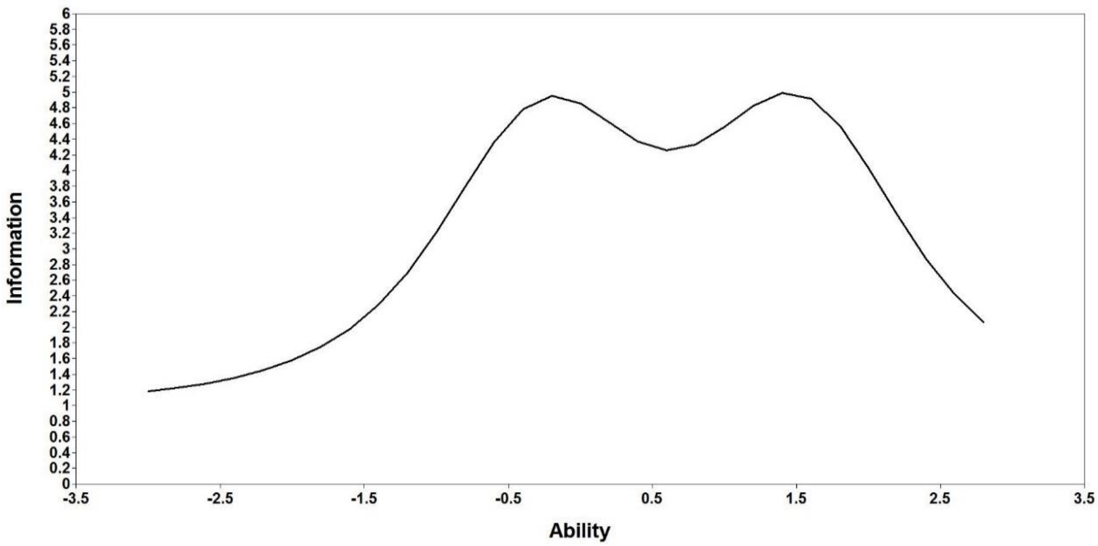


Figure 4.7
Test Information Function for the KFCC Scenario

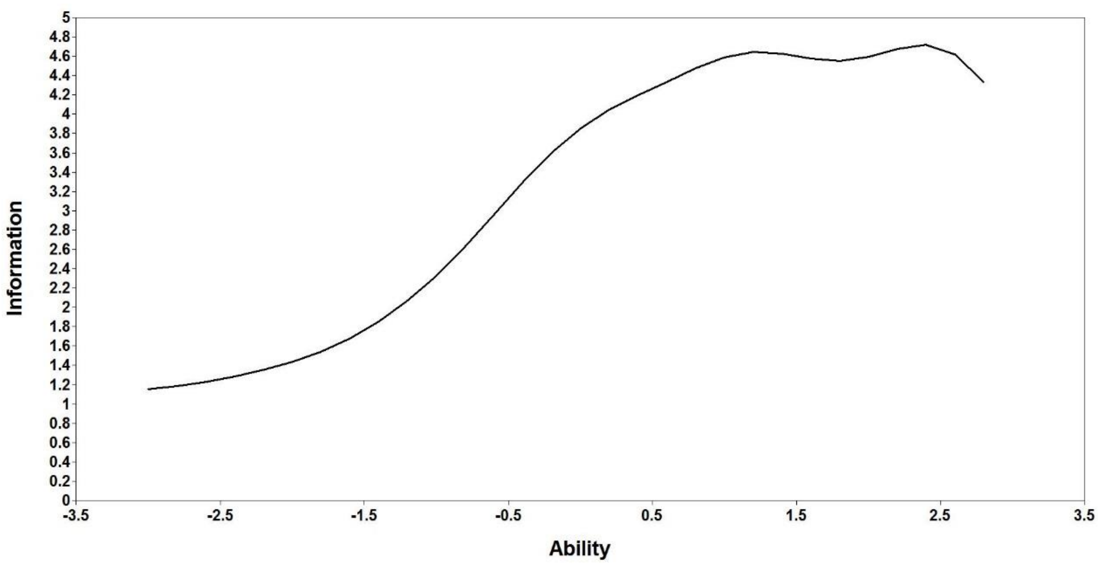


Figure 4.8
Test Information Function for the BBD Scenario

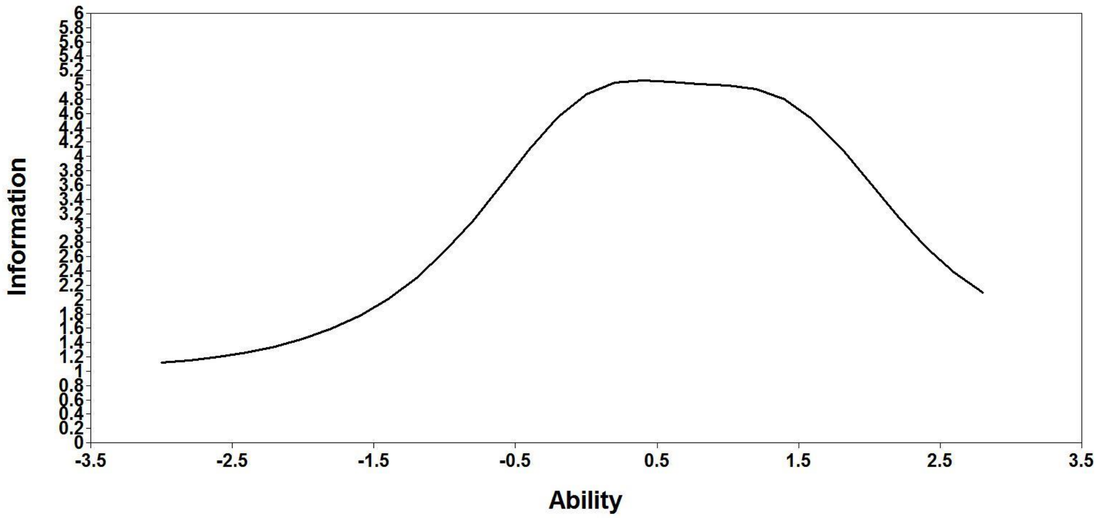


Figure 4.9
Test Information Function for the OTO Scenario

Tables 4.11, 4.12 and 4.13 provide a summary of the findings from each scenario. Table 4.11 shows the summary of findings for the KFCC scenario. Given the information described in Table 4.11, Questions 2 and 4 should be revised despite the fact that they had strong factor loadings, because they did not provide much information about the students' ability to think critically about sustainability on the scenario and students engaged in low-level critical thinking to respond these questions. Question 1 needs revision because despite the fact that it provided low information and had a weak factor loading it assessed students' higher order thinking skills. Questions 3 and 5 should be retained for further use given their high factor loadings, high information, and assessment of students' higher-order thinking skills. Question 6 has great potential to be a very good item for scale in lieu of revisions to its grading criteria that have affected the reliability of the scoring. Question 6 possesses measures students' metacognitive skills, has a high factor loading (0.64) and high slope values (1.47). However, one concern remains about the KFCC scenario in that the test information is bimodal making it difficult to select items for the scale due to fact that it provides the most information at two latent ability levels. In light of this, strong consideration it should be given to either replacing or rewriting the KFCC scenario.

Table 4.13
Summary of Findings for the KFCC Scenario in the CTSS

Question	Information	*r*	Loadings	Difficulty	
				T1	T2
Describe the sustainability issues and challenges	Low	Low	Weak	D	VD
Identify additional information needed to thoroughly consider the sustainability issues	Medium	High	Strong	MD	D
Recommend a strategy for resolving the sustainability issue. How or why would this strategy resolve the sustainability issue?	High	High	Strong	E	VD
Who are the key stakeholders? In what ways are their interest aligned or in competition?	Medium	High	Strong	VE	MD
Describe the implications and consequences of your recommendation.	High	High	Strong	E	VD
Describe how your own knowledge, perspectives, and opinions influence interpretation of this case study.	Medium	High	Strong	MD	VD

Note. Information < 1 = Low, 1 < Information < 2 = Medium, Information > 2.0 = High.0Low indicates Correlation (*r*) < 0.30; weak factor loading = loading < 0.32; Difficulty: Very Easy (VE) = T1/T2 < -1.0, Easy (E) = -1.0 < T1/T2 < 0.0, Moderate Difficulty (MD) = 1.0 < T1/T2 > 0.0 Difficult (D) = T1/T2 > 1.0, Very Difficult (VD) = T1/T2 > 2.0

CHAPTER 5. DISCUSSION

There is need for a psychometrically sound instrument to measure undergraduate engineering students' ability to think critically about sustainability. It is evident from the literature and the expert interviews that sustainability is a multifaceted construct that requires persons to think critically in order to discover or develop ways to create a sustainable world.

Engineering students need to be able to think critically about a wide range of sustainability issues, and not only those related to the environment. If the world is to achieve sustainability, to ensure that the earth can maintain life in all its forms, that current and future generations can survive and thrive, and that equality and social justice are not just for the privileged few but a right for everyone then engineering—nay higher education—needs to produce graduates who are capable of rising to the challenge of creating a sustainable world. Thus, the primary objectives of this study was to understand how sustainability experts were similar or different in their conceptualizations of sustainability and to evaluate the extent to which the questions on the Critical Thinking about Sustainability Scale (CTSS) measured undergraduate engineering students' ability to think critically about sustainability.

This section first presents the major findings of each research question and implication of these findings, beginning with experts' conceptualization of sustainability, followed by the results of the evaluation of the construct validity and item quality of the CTSS. The discussion closes with some of the limitations of the study and future directions. As the part of the discussion of future direction of the study on the CTSS, Tables 5.1, 5.2 and 5.3 provides summaries of the recommendations for revising the CTSS for each scenario.

5.1 Experts' Conceptualization of Sustainability

The results of the expert interviews showed a number of similarities and differences in experts' conceptualization of sustainability. Most notably, all the experts shared the sentiment that sustainability was a difficult topic to conceptualize. This view was consistent with the literature where researchers cited the lack of a clear understanding of sustainability and sustainable development hampers the integration of sustainability principles in university curricula (Thomas, 2009). Another point to note is that the experts all came from diverse backgrounds and had their own views about sustainability. It has been noted in the literature that researchers and stakeholders (e.g., Littig & Grießler, 2005; Thomas, 2009; WCED, 1987) have their own conceptualizations of sustainability and what a sustainable society looks like. For example, Chichilnisky (1997) argues against the practice of discounting intergenerational equity that is key aspect of sustainability often neglected in economic definitions of sustainability (Pezzey & Toman, 2001; Toman, 2006). One distinct difference between the experts' views on sustainability and the literature is the lack of discussion about the economic aspect of sustainability. For example, in the expert interviews, Expert C disregarded the economic aspect of sustainability as it came a distant second to ecological sustainability. In contrast, Expert B treated all aspects of sustainability of equal importance. These experts' views stand in contrast to literature where the concept of economic sustainability addressed explicitly by researchers. These various views about sustainability are what makes the conceptualization sustainability and its integration into higher education difficult (Thomas, 2009).

Equity was another concept that all experts expressed as important to defining sustainability. Particularly, they stressed equity in the distribution of the earth's natural resources and the equality of opportunities for people and societies thrive and survive. This notion of equity was noted in the literature where researchers espoused the idea that concerns about environmental sustainability relate to issues of social justice, gender-equality, and political participation (Becker et al., 1999). In addition, all the experts espoused the idea of sustainability having a time component was expressed by the experts in the interviews. This view was consistent with Waas (2011) claims that sustainability is the best way to address the interrelated environmental and societal problems that affect both current and future generations. The experts' conceptualization of sustainability provided a solid foundation for defining sustainability in higher education to apply for the development the CTSS. As a result, the sustainability was conceptualized as "the earth's resources are utilized in an equitable and judicious manner; current and future generations are able to survive and thrive economically and socially; no harm is brought to the earth's ability to sustain life" in the current study.

However, given the myriad definitions of sustainability, continued discussion on achieving consensus about what sustainability means is needed in future studies. At the institutional level, the shared definition would be useful for the application of sustainability issues in higher educations. In particular, sustainability experts and the researchers (Chau, 2007; Thomas, 2009) expressed concern about teaching sustainability principles in higher education. As Moore (2005) notes, incorporating sustainability into higher education is essential for students to understand how their everyday actions can affect the relationship among environmental, economic, and social issues and to influence their actions as local and global citizens.

I propose the integration of safe minimum standards for sustainability (SMS; Ciriacy-Wantrup, 1952) into policy decisions to set a socially derived dividing line between moral obligations to preserve and enhance natural resources and the free use of resources trade-offs (Toman, 2006) as means of addressing key sustainability issues. Next, not explicitly mentioned in the literature were gender differences in students' approaches to sustainability issues, while Expert B mentioned distinct differences between males and females on their engagement and approach to sustainability issues in the classroom. Researchers have attempted to measure students' knowledge and affective reaction toward sustainability and sustainable development with a quantitative scales (Azapagic, et al., 2005; Kagawa, 2007).

For example, Kagawa (2007) found that a higher proportion of male students favored technological solutions as a means of achieving a better future society while higher proportion of female students opted for the formation of local economies as the way to a better society. However, in this study, the gender differences were limited to students' personal views about the future of society and not the other sustainability issues addressed in the study. Similarly, Azapagic et al. (2005) developed an international survey to investigate engineering students' knowledge of sustainable development but did not explore gender differences in students' responses. Expert C expressed the view that sustainability is usually described in feminine rather and masculine terms and thus addressing sustainability issues requires addressing the male/female dichotomy. Gender related issues concerning learning and teaching of sustainability concepts has very important future implications for the field of engineering as females' unique contribution to the field may lost due to lack of representation in STEM fields (Halpern, 2007).

Finally, experts consistently reported that students' view of sustainability is unidimensional and they tend to focus the ecological aspect of sustainability. Qualitative students' responses on Question 7 (describe what sustainability means to them) are consistent with expert's claim. Only four students out of 151students made references to the economic and social aspects of sustainability. This finding also supported the idea by the experts that students' conceptualized sustainability in distinctively ecological ways. This information was very useful for designing the CTSS that capture three elements in sustainability in questions as the previous scale tend to focus on ecological aspect of sustainability (Azapagic et al., 2005; Kagawa, 2007)

The experts' views of sustainability also suggest some insights into the implementation of sustainability issues in engineering. There is need for more interdisciplinary work in engineering to help extend the conceptualization of sustainability to include the economic and social aspects of sustainability to more explicitly (Toman, 2001). One possible way to achieve this will be for faculty with differing views on ecological and economic sustainability to engage in interdisciplinary research. For example, economists may expand their evaluation of resources to include the function and value of ecosystems as a whole and to prioritize socially optimal allocation of resources that account for inter- and intra-generational concerns over economically efficient allocations of resources (Toman, 2001). Ecologists may help by disseminating ecological information in a form that economists can use when allocating values to resources and they should recognize that human behavior and social decision-making processes are just as complex as ecological processes.

These cross-disciplinary studies can then be adapted and incorporated into the engineering curricula in the form of design projects and scenario-based tasks that are currently widely used in engineering. In fact, this will help not only engineering students, but also students in other disciplines get a greater understanding of sustainability as an interwoven phenomenon that affects every area of human existence. In this way, students in higher education will have a greater appreciation of sustainability and will be more active and engaged in learning and incorporating sustainability principles both in school and in the workplace in the future.

5.2 Evaluation of CTSS Construct Validity and Quality

Overall, a series of psychometric analyses per scenario provided evidence that the CTSS measures a single construct that represented students' ability to think critically about sustainability. Thus, the study showed supportive evidence for the construct validity of CTSS. The analyses of the item and scale quality indicated that the scenarios provide limited to adequate information about students' ability to think critically about sustainability and need revision to improve the quality of the scale. Across the three scenarios, questions measuring students' ability to analyze information (Question 1) and their metacognitive skills (Question 6) were the most difficult for students while questions measuring their ability to identify information (Question 2) and to identify and analyze information (Question4) were the easiest for the students in the study. Questions

3 and 5, measuring students' problem solving and evaluation skills, respectively, are identified as adequate through the analyses.

More importantly, the study identified the necessary revisions at the question level to enhance the reliability of the CTSS for subsequent use in practice. Therefore, I summarize and discuss the results of the psychometric evaluation of each scenario for future revisions. For the KFCC scenario, Cronbach's Alpha (α) was 0.69 (Table 4.5), indicating less-than adequate to adequate scale reliability based on aforementioned reliability criteria of 0.70 (Nunnally, 1978). Therefore, including more items in each scenario with similar quality or administering the scale to a larger group of students with diverse ability levels may provide improved reliability evidence of the scale (Gregory, 2007). However, due to the use of open-ended questions extra precautions should be taken to ensure that the scoring rubric is clear and the questions are well developed to enhance the reliability and validity of the scale.

Question 1 was one of the more difficult questions in the scenario. It generally had the lowest item-total correlations (0.19), factor loading (0.30), and item information (0.58). The wording of the question may be the main cause of its poor psychometric properties. An option for improving this question involves rewording the item stem by using either the word "issue" or "challenge" which removes "and" and "or" that makes the question unnecessarily complicated and thereby improving the quality of responses to the items.

Like Question 1, Question 6 was a difficult for students and needs revision. This question evaluates students' metacognitive ability, which is an important critical thinking skill. Overall, this this was a fair question since it had strong factor loading (0.64), provided an adequate amount of information (1.47), and had a very good item-total correlation (0.44). However, the main issue lay not with the question itself, but its scoring. It is evident from Table 4.3 that Question 6 had the lowest inter-rater agreement indicating problems with the scoring. This was the result of students having such a diverse range of experiences that influenced their responses to the question, making scoring using the current criteria difficult. For example, students reported that their responses to the question were influenced experiences such as the courses they took, their family experiences, travel experiences, extra curricula experiences, experiences with peers, and the media. Therefore, the current scoring rubric might not capture the full dimension of possible student responses to this item so revising the scoring criteria is necessary.

Another possible option to improve Question 6 would be to increase the number of categories in the scoring rubric. However, researchers have advocated against the use of more than three categories because it is difficult for individuals to make sense of information from rubrics with more than three categories (Bresciani, Zelna, & Anderson, 2004). Therefore, further evaluation of written responses gathered in the current study would be necessary to determine the optimal number of the scoring categories and rubric.

Finally, the wording of the question could be improved. For example, Question 6 asks students to describe how your own knowledge, perspectives, and opinions influence interpretation of this case study. The question could ask instead, “How have your experiences influenced your responses to the questions?”, or “In what ways have your experiences influenced your responses to the questions?” Then responses for each category can be more specific about the criteria used to assign scores. Therefore, a *proficient* rating (i.e., score of 3) criteria in Questions 6 is, “the student describes in detail how their own knowledge, perspectives, and opinions influence their interpretation of this case study”. An optional rewording can be, “the students describe in detail how *at least one specific activity or experience* influenced their responses to the question”. Adding the requirement of at least one specific activity or experience provides a “starting point” for evaluating the level of detail about responses.

Following the *proficient* rating, the criterion for obtaining the *emerging* rating (i.e., score of 2) on Question 6 can be changed from, “the students provide some detail about how their own knowledge, perspectives and opinions influence their interpretation of this case study” to “the students describe in some detail how at least one specific activity influenced their responses to the question”. The *developing* rating (i.e., score of 1) can be reworded from, “the students *do not* describe in detail how their own knowledge, perspectives, and opinions influence their interpretation of this case study” to “the students *fail* to describe how at least one specific activity influenced their responses to the question”.

Questions 2 and 4 in the KFCC had strong factor loadings 0.56 and 0.58, respectively, and generally had the largest proportion of students obtaining 3 points. The major problem with Questions 2 and 4 were that they were unable to discriminate between students with high and low on the latent trait and thus they provided very little information about the CTSS. Evaluation of the Question 2 responses also suggest the possibility of replacing the item format to the closed-end form (i.e., multiple items) because students tended to respond to the question using bullet points or numerical lists. This response style also implies that the students were not likely to be engaged in any form of critical thinking with the question. A possible revision on wording for Question 2 may be, “how can you assess the importance of the sustainability issues described in Question 1?” This change would likely to increase students’ engagement in the process of evaluating information rather than just identifying information from the text.

Question 4 asked students to, “identify all the stakeholders involved and how their interests are aligned or in competition”. An alternative way to phrase this question would be, “how can you identify the key stakeholders in this scenario?” that would require students to go beyond the simple process of identifying the stakeholders to explaining how they came to that conclusion. Then, for the scoring rubric, they would receive a *proficient* score if, “the student describes how they identified at least 3 stakeholders”. They would receive an *emerging* score if, “the student describes how they identified at least 2 two stakeholders”. The student would receive a *developing* score if, “the student describes how they identified at least one stakeholder”.

Another point to consider on revising Question 4 is that while the scenario as whole provided adequate information about students’ ability to think critically about sustainability, the TIF showed a distinct bimodal shape. From Figure 4.7 it is easy to see that the TIF for the KFCC scenario has two distinctive peaks that provide the most accurate information at approximately latent trait level -0.2 and latent trait 1.5. Thus, making it difficult to determine the latent ability level that maximizes the amount of information on the CTSS (Reckase, 2009), although this is often observed in the ordered rating scale with two categories (Baker, 2009). However, one should see improvement in the shape of the TIF with the inclusion of the revised questions.

Compared to the first scenario, Questions in the Broken Bulb Dreams (BBD) scenario generally function better in terms of measuring critical thinking about sustainability. Question 1 the BBD scenario was the second least precise item of the six questions in the scenario, and had the second lowest item information of the scenario (1.01) as seen on Table 4.12. Given that the same questions are used in each scenario, the differences in the factor loadings, item-total correlations and item information may be attributed to differences across scenarios may be the product of the content of the scenario. For example, students may have been better able to respond to Question 1 of the BBD scenario because they may have found it easier to describe the sustainability issues and challenges of the scenario. However, apart from the differences in the content, Question 1 in the BBD scenario suffers from the same major problem as Question 1 of the KFCC scenario—poor item wording. Therefore, as recommended for Question 1 of the KFCC scenario the wording of the BBD scenario should use either the word “issue” or “challenge” which removes “and” and “or” that makes the Question unnecessarily complicated.

Unlike Question 6 of the KFCC scenario, Question 6 of the BBD scenario had a higher factor loading (0.82), higher item-total correlation (0.48), and higher information (2.40). Again, given the fact the question is the same for both scenarios the content of the scenario may have been the contributing factor to the differences in the factor loadings. Question 6 of the BBD scenario suffers from the same issue of the scoring rubric wording therefore the changes outlined for the KFCC scenario are recommended.

Question 2 of the BBD scenario had a slightly higher factor loading (0.62), item-total correlation (0.42), and item information (1.26). For this question, it seems that the content of the scenario did not significantly affect student responses. Since Question 2 was the same for both scenarios, the same problem of low information about the CTSS and measuring low-level cognitive skills persisted for Question 2 in the BBD scenario. Thus, the recommended revisions for Q2 of the KFCC scenario should be adopted.

Question 4 of BBD scenario did not function as well as its counterpart in the KFCC scenario. This question had a lower factor loading (0.47), lower item-total correlation (0.31), and lower information (0.93). Students' responses to this question may have been affected by differences in the content of the scenarios. It also suffers from the same problems of not assessing higher-order thinking skills of students. Therefore, revisions to Question 4 should follow the recommendations from the KFCC scenario. The TIF for the BBD scenario had a slightly bimodal shape and provided a little less information than the TIF of the KFCC scenario. However, the maximum inform obtained from the TIF of the BBD scenario occurred at higher ability levels. Therefore, the BBD scenario provided more information at a higher ability levels than the KFCC scenario and may be more suitable for higher ability students.

The pattern of results for the OTO scenario was more similar to those of the KFCC scenario, than the BBD scenario. Question 1 of the OTO scenario had the lowest factor loading (0.45), and the lowest information (1.01). However, unlike the other scenarios, Question 1 of the OTO was so difficult that the program was unable to estimate the second threshold for the data. Like the BBD scenario, the content of the scenarios may have affected their responses and the wording of the question needs revision. The revision of this question should follow the recommendations outline for the KFCC scenario. For Question 6, the factor loading was high (0.71) and provides a moderate amount of information similar to Question 6 of the KFCC scenario indicating that the content of the scenario did not unduly affect students responses to the question. Question 6 again, suffers from the problem of the scoring rubric that can be improved by changing the question stem and scoring rubric in manner described for the KFCC scenario.

Questions 2 and 4 of the OTO had strong factor loadings 0.55 and 0.63, as well as the second lowest item-total correlations 0.41 and 0.46 and the second lowest amount of information (i.e., 1.10 and 1.42, respectively) and generally had the largest proportion of students obtaining 3 points. Except for Question 4 of the BBD scenario, the questions did not seem to be affected to great extent by the content of the scenarios. As the other scenarios, the major problem with Questions 2 and 4 were low information provided about the CTSS and measuring low-level cognitive skills. Like the other two scenarios, Question 2 can be replaced with the question stem described in the first scenario. Similarly, Question 4 can be revised using the aforementioned recommendations. Of the three TIFs the function for the OTO shows the most symmetrical shape and providing information at an above average ability making it more suitable to a general population of students.

Finally, consideration should be given to the continued use of open-ended questions. The open-ended questions were fairly challenging and time consuming to score, particularly Question 6, whose results may have been affected the most by the scoring. However, it is recommended that the open-ended question format be retained in further iterations of the study. Despite the difficulties in scoring and time-intensiveness associated with scoring open-ended questions the richness of information obtained provided valuable insight into the students' ability to think critically about sustainability. The quality of information obtained by more objective and easier to score test formats would be less due to lack of opportunity to provide detailed justifications for their answer choices and the freedom to provide more than one correct answer (Renaud & Murray, 2004). Therefore, the ability to provide more than one correct answer and to justify ones actions are key skills needed by students to be able to deal with the complex topic of sustainability that are not provided by multiple choice type questions (Renaud & Murray, 2004). Therefore, the use of such scales will limit the ability of students to utilize the full range of their critical thinking skills to answer questions about sustainability (Ennis, 1993; Halpern, 2003; McPeck, 1981). Tables 5.1 -5.3 summarized the suggested changes on the CTSS for the future validation.

Table 5.1
Decision Criteria for Removing or Revising Questions on the KFCC Scenario

Question	Quality	Suggested Revision	Recommendations for Revisions
1	Poor	Revise	• Revise question wording
2	Poor	Revise	• Revise scoring rubric
3	Good	Retain as is	• No revisions required
4	Poor	Revise	• Revise item wording • Revise scoring rubric
5	Good	Retain as is	• No revisions required
6	Fair	Revise	• Revise question wording • Revise scoring rubric

Table 5.2
Decision Criteria for Removing or Revising Questions on the BBD Scenario

Question	Quality	Suggested Revision	Recommendations for Revisions
1	Poor	Revise	• Revise question wording
2	Poor	Revise	• Revise scoring rubric
3	Good	Retain as is	• No revisions required
4	Poor	Revise	• Revise item wording • Revise scoring rubric
5	Good	Retain as is	• No revisions required
6	Fair	Revise	• Revise question wording • Revise scoring rubric

Table 5.3
Decision Criteria for Removing or Revising Questions on the OTO Scenario

Question	Quality	Suggested Revision	Recommendations for Revisions
1	Poor	Revise	• Revise question wording
2	Poor	Revise	• Revise scoring rubric
3	Good	Retain as is	• No revisions required
4	Poor	Revise	• Revise item wording • Revise scoring rubric
5	Good	Retain as is	• No revisions required
6	Fair	Revise	• Revise question wording • Revise scoring rubric

5.3 Limitations and Future Research

In summary, the CTSS shows sound psychometric properties as a pilot scale to measure undergraduate engineering students' ability to think critically about sustainability. The CTSS holds much promise for future use in the field of engineering and possibly other disciplines. Since the CTSS is best for estimating the scores of students whose latent trait occur near the average to above-average difficultly, the most appropriate use of the scale would be to distinguish respondents who are high on the ability to think critically about sustainability and students who are low on the ability to think critically about sustainability (Embretson & Reise, 2000).

Researchers may use the CTSS with future revisions in a pre-post test design to diagnose students' ability to think critically about sustainability that will allow researchers to develop targeted interventions to improve their students' learning. Researchers can use the CTSS provide their students with valuable feedback about their critical thinking processes that may motivate them to become better critical thinkers (Ennis, 1993). Researchers may also use the CTSS for evaluative purposes as means to inform instructors about how successful they were in teaching their students to think critically about sustainability (Ennis, 1993).

While the study achieved its intended purpose, there are limitations to the study that can inform further investigations in the future. The major concern was the sample size of the study (N = 151). With regard to exploratory factor analysis, the generally recommended sample size of at least 200 participants, to obtain meaningful results were not met (Floyd & Widaman, 1995; Gorsuch, 1983). A sample size of approximately 200 students would have produced more stable estimates. However, the statistically significant and moderate to strong factor loadings on the questions may have mitigated the effect of a lack of power for EFA in the study. With regard to the 2PL graded-response model the study did not meet the recommended sample size of 250-500 participants to produce accurate parameter estimates" (Reeve & Fayers, 2005). The ideal next step for this study would be to conduct another study with a sample of 250-500 students to cross-validated the revised CTSS with the edited items and additional items with EFA and with the 2PL graded-response model to evaluate the effect of the item revisions on the latent structure of the CTSS. With a sufficiently large sample (e.g., 750), as the final stage of the CTSS development, CFA can be used to validate the structure of the CTSS. Additionally, larger sample sizes will enable to test the factor structure invariance between males and females and between scenarios as means of empirically evaluating group differences.

Another limitation related to the sample of the study was the general characteristics of the site and subjects. In this study, the data were collected using a purposive sampling method restricted to the population of engineering students and thus the results are only confined to undergraduate engineering students and affecting the external validity of the scale (Gay, Mills, & Airasian, 2006). Although the original aim to develop this scale is for the use with undergraduate engineering students, it may be ideal to use the scale with other undergraduate population. Thus, it may be necessary to validate the revised CTSS with diverse student population. This will also provide a great opportunity to gain a deeper insight of how the CTSS measures students' ability to think critically about sustainability and aid with the scale's development.

Finally, this study evaluated the dimensionality of the data for the scenarios separately to avoid confounding the unique latent structures of the scenarios. Future studies should incorporate multi-group invariance testing in order to ascertain whether the factor loadings are invariant across the scenarios. Like the recommendations for future EFA and IRT analyses invariance testing of the CTSS should involve larger samples that can ensure reliable estimates of the models.

5.4 Conclusion

To date, little research exists on this concept and the findings of this research can stimulate more investigations into the topic and fill a major research gap in the field of students' ability to think critically about sustainability. This study, despite its shortcomings, shows promise for the development and validation of the CTSS. Thus far, the development and validation of the CTSS is the first of its kind in the field of sustainability and the experts all agree that there is a great need for a scale capable of measuring students' ability to think critically about sustainability.

More specifically, researchers can use the results of the current psychometric evaluation to further improve the scale for future use. Once the scale is fully developed, researchers and educators can have a medium to use for research and in practice relevant to critical thinking about sustainability. This study can serve as the starting point for the construction of a database of empirically validated scenarios that researchers, educators, and administrators can use in their classes to assess students' ability to think critically about sustainability. Finally, although the CTSS was utilized in an engineering context in the current study, the scenarios that comprise the scale are discipline general such that the tool can be utilized in a diverse population of students. Therefore, the evaluation of the

CTSS in this study will contribute to advance research on understanding and promoting undergraduate engineering students' ability to think critically about sustainability and ultimately to students in a broader spectrum of disciplines.

LIST OF REFERENCES

LIST OF REFERENCES

Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch Rating Model and the Disordered Threshold Controversy. *Educational and Psychological Measurement, 72*(4), 547-573. doi: 10.1177/0013164411432166

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. Tesol Quarterly, *31*(1), 71-94.

Azapagic, A., Perdan, S., & Shallcross, D. (2005). How much do engineering students know about sustainable development? The findings of an international survey and possible implications for the engineering curriculum. *European Journal of Engineering Education*, *30*(1), 1-19. doi:10.1080/03043790512331313804

Badger, E., & B. Thomas, (1992). Open-ended questions in reading. *Practical Assessment, Res. and Evaluation*, *3*(4). Retrieved from http://pareonline.net/getvn.asp?v=3&n=4

Bagozzi, R., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74-94. doi: 10.1007/bf02723327

Bailin, S. (2002). Critical thinking and science education. *Science & Education, 11*(4), 361–375. doi:10.1023/A:1016042608621

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies, 31*(3), 285–302. doi: 10.1080/002202799183133

Baker, Frank. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

Becker, E., Jahn, T. and Stieß, I. (1999). 'Exploring uncommon ground: sustainability and the social sciences'. In Becker, E. and Jahn, T. (Eds.), *Sustainability and the Social Sciences, A Cross-Disciplinary Approach Integrating Environmental Considerations into Theoretical Reorientation* (pp.1–22). London: Zed Books.

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychological Bulletin, 107*(2), 238-246. doi: 10.1037/0033-2909.107.2.238

Bethlehem, J., & Biffignandi, S. (2012). *Handbook of Web surveys*. New York: Wiley.

Bloom, B. S., Jr. (1956). *Taxonomy of educational objectives: handbook I. The cognitive domain.* New York: McKay.

Bresciani, M. J., Zelna, C. L., & Anderson, J. A. (2004). Assessing student learning and development. *A handbook for practitioners. United States: NASPA.*

Brookfield, S. D. (1987). *Developing critical thinkers: Challenging adults to explore alternative ways of thinking and acting.* San Francisco: Jossey-Bass Publishers.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*: Guilford Press

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis. In Grimm, L. G., & Yarnold, P. R. (Eds), *Reading and understanding multivariate statistics* (pp. 99-136). Washington, DC, US: American Psychological Association.

Case, R. (2005). Moving critical thinking to the main stage. *Education Canada, 45*(2), 45–49.

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research, 1*(2), 245-276. doi: 10.1207/s15327906mbr0102_10

Chau, K. W. (2007). Incorporation of sustainability concepts into a civil engineering curriculum. *Journal of Professional Issues in Engineering Education and Practice, 133*(3), 188-191.

Chichilnisky, G. (1997). What is sustainable development?(Special Issue: Defining Sustainability). *Land Economics, 73*(4), 467-491.

Christian, Leah Melani, Don A. Dillman, and Jolene D. Smyth. 2007. "Helping Respondents Get it Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys." *Public Opinion Quarterly*, 71, 113-25.

Ciriacy-Wantrup, Sigfried von. (1952). Resource conservation. Berkeley: University of California Press.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological assessment*, *7*(3), 309-319. doi: 10.1037/1040-3590.7.3.309

Clugston, R. M., & Calder, W. (1999). Critical dimensions of sustainability in higher education. *Sustainability and university life*, *5*, 31-46.

Cook, K. F., Dodd, B. G., & Fitzpatrick, S. J. (1999). A Comparison of Three Polytomous Item Response Theory Models in the Context of Testlet Scoring. *Journal of Outcome Measurement*, *3*(1), 1-20.

Costanza, R., & Daly, H. E. (1987). Toward an ecological economics.*Ecological Modelling*, *38*(1), 1-7. doi:10.1016/0304-3800(87)90041-X

Cortese, A. D. (2003). The critical role of higher education in creating a sustainable future. *Planning for higher education*, *31*(3), 15-22.

Costa, S., & Scoble, M. (2006). An interdisciplinary approach to integrating sustainability into mining engineering education and research. *Journal of Cleaner Production*, *14*(3), 366-373. doi:10.1016/j.jclepro.2004.01.011

Council for Aid to Education (2015). Collegiate Learning Assessment [Assessment]. Published instrument. Retrieved from http://cae.org/participating-institutions/cla-overview

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. doi: 10.1037/h0040957

Cuthill, M. (2010). Strengthening the ‘social’in sustainable development: Developing a conceptual framework for social sustainability in a rapid urban growth region in Australia. *Sustainable Development*, *18*(6), 362-373. doi:10.1002/sd.397

Daily, G. C., & Ehrlich, P. R. (1992). Population, sustainability, and Earth's carrying capacity. *BioScience*, *42*(10), 761-771.

Daily, G. C., & Ehrlich, P. R. (1996). Socioeconomic equity, sustainability, and Earth's carrying capacity. *Ecological Applications*, *6*(4), 991-1001. doi:10.2307/2269582

Davidson, C. I., Hendrickson, C. T., Matthews, H. S., Bridges, M. W., Allen, D. T., Murphy, C. F., ... & Austin, S. (2010). Preparing future engineers for challenges of the 21st century: Sustainable engineering. *Journal of Cleaner Production*, *18*(7), 698-701. doi:10.1016/j.jclepro.2009.12.021

Davino, C. (2013). *Survey Data Collection and Integration [delta]*. Berlin, Heidelberg: Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (Vol. 26). Sage Publications.

Dewey, J. (1910). *How we think.* Boston: D. C. Heath and Company Publishers.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: the total design method.* New Jersey: John Wiley & Sons.

Duflo, E. P., & Pande, R. (2007). Dams. *Quarterly Journal of Economics*, *122*(2), 601-646.

Du, X., Su, L., & Liu, J. (2013). Developing sustainability curricula using the PBL method in a Chinese context. *Journal of Cleaner Production*, *61*, 80-88. doi: 10.1016/j.jclepro.2013.01.012

Edgerton, R. (1991). National standards are coming!...National standards are coming! *AAHE Bulletin.* 44(4), 8-12.

El Serafy, S. (1991). The environment as capital. In Costanza, R. (Ed.) *Ecological economics: The science and management of sustainability* (pp. 168-171). New York: Colombia University Press

Embretson, S. E. & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, N.J.: Mahwah, N.J. : L. Erlbaum Associates.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, *43*(2), 44–48.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, *18*(3), 4–10.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (Executive summary of "The Delphi Report")* Millbrae, CA: California Academic Press.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, *20*(1). 61-84.

Facione, P. A. & Facione, N. C. (1992). *The California Critical Thinking Dispositions Inventory (CCTDI); and the CCTDI Test manual*. Millbrae, CA: California Academic Press.

Facione, P. A., Sánchez, C. A., Facione, N. C., & Gainen, J. (1995). The disposition toward critical thinking. *The Journal of General Education*, *44*(1). 1-25.

Ferrer-Balas, D., Lozano, R., Huisingh, D., Buckland, H., Ysern, P., & Zilahy, G. (2010). Going beyond the rhetoric: system-wide changes in universities for sustainable societies. *Journal of Cleaner Production, 18*(7), 607-610. doi: 10.1016/j.jclepro.2009.12.009

Foy, G. (1990). Economic sustainability and the preservation of environmental assets. *Environmental Management*, *14*(6), 771-778. doi: 10.1007/BF02394171

Fischer-Kowalski, M., & Haberl, H. (1993). Metabolism and colonization. Modes of production and the physical exchange between societies and nature. *Innovation: The European Journal of Social Science Research*, *6*(4), 415-442. doi: 10.1080/13511610.1993.9968370

Fricker, A. (1998). Measuring up to sustainability. *Futures*, *30*(4), 367-375. doi: 10.1016/S0016-3287(98)00041-X

Gary J. & Esquivel, Shelley L. (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational Research.*Practical Assessment, Research & Evaluation*, *18*(6), 1-13

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational research: Competencies for analysis and applications* (8th. ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Ginther, D., & Kahn, S. (2006). Does Science Promote Women? Evidence from Academia 1973-2001. Cambridge, Mass., USA.

Glaser, E. M. (1941). *An experiment in the development of critical thinking.* New York: Teachers College of Columbia University, Bureau of Publications.

Goodland, R. (1995). The concept of environmental sustainability. *Annual Review of Ecology and Systematics*, *26*, 1-24.

Goodland, R., & Daly, H. (1996). Environmental sustainability: universal and non-negotiable. *Ecological Applications*, *6*(4), 1002-1017. doi: 10.2307/2269583

Gutierrez-Martin, F., & Dahab, M. F. (1998). Issues of sustainability and pollution prevention in environmental engineering education. *Water science and technology*, *38*(11), 271-278.

Haberman S. J. (2006). *An elementary test of the normal 2PL model against the normal 3PL alternative.* (ETS Research Rep. RR-06-14). Princeton, NJ: ETS.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist*, *53*(4), 449–455.

Halpern, D. F. (2001). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, *50*, 270–286.

Halpern, D. F. (2003). The “how” and “why” of critical thinking assessment. In Fasco, Jr. D. (Ed.) *Critical thinking and reasoning. Current research, theory, and practice* (pp. 331-354). New Jersey: Hampton Press, Inc.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. [Article]. *Psychological Science in the Public Interest (Wiley-Blackwell), 8*(1), 1-51. doi: 10.1111/j.1529-1006.2007.00032.x

Hans-Boeckler-Foundation. (2001), *Pathways towards a sustainable future*. Setzkasten: Düsseldorf.

Hardin, G. (1993). *Living within limits*. New York: Oxford University Press.

Hartwick, J. M. (1977). Intergenerational equity and the investing of rents from exhaustible resources. *American Economic Review, 67*, 972-974.

Haladyna, T. M. (1997). *Writing Test Items To Evaluate Higher Order Thinking*. Massachusetts : Allyn & Bacon.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, *15*(3), 309-333. doi: 10.1207/S15324818AME1503_5.

Hays, D. R., Morales, S. L., & Reise, P. S. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical care*, *38*(9 Suppl II), II-28-II-42.

Holden, M., Elverum, D., Nesbit, S., Robinson, J., Yen, D., & Moore, J. (2008). Learning teaching in the sustainability classroom. *Ecological Economics, 64*(3), 521-533. doi: http://dx.doi.org/10.1016/j.ecolecon.2007.09.007

Husain, H., Bais, B., Hussain, A., & Samad, S. A. (2012). How to Construct Open Ended Questions. *Procedia - Social and Behavioral Sciences*, *60*(0), 456-462. doi: http://dx.doi.org/10.1016/j.sbspro.2012.09.406

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55

Inquirium (2005-2011). Inqscribe [Computer software]. Retrived from http://www.inqscribe.com/

Insight Assessment (2013). California Critical Thinking Skills Test [Measurement Instrument]. Published instrument. Retrieved from http://www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/California-Critical-Thinking-Skills-Test-CCTSS

Jöreskog, K.G. & Sörbom, D. (1986). *LISREL VI, analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*: Mooresville, Ind.: Scientific Software, Inc.

Kagawa, F. (2007). Dissonance in students' perceptions of sustainable development and sustainability: Implications for curriculum change. *International Journal of Sustainability in Higher Education, 8*(3), 317-338. doi: 10.1108/14676370710817174

Kelman, S. (1981). *What price incentives?: economists and the environment*. Boston: Auburn House.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.. ed.). New York: New York : Guilford Press.

Kline, T. (2005). *Psychological testing : a practical approach to design and evaluation*. Thousand Oaks, Calif.: Thousand Oaks, Calif. : Sage Publications.

Ku, K. Y. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking skills and creativity*, *4*(1), 70-76. doi:10.1016/j.tsc.2009.02.001

Kulonda, D.J. (2001). Case learning methodology in operations engineering. *Journal of Engineering Education*, *90*(3), 299–303. doi: 10.1002/j.2168-9830.2001.tb00608.x

Kurfiss, J. G. (1998). *Critical thinking: Theory, research, practice, and possibilities.* ASHE-ERIC Higher Education Report No. 2, 1988. Washington: Association for the Study of Higher Education.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports. 6*, 40-41.

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into practice*, *32*(3), 131-137. doi:10.1080/00405849309543588

Lincoln, YS. & Guba, EG. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications

Lipman, M. (1988). Critical thinking—What can it be? *Educational Leadership, 46*(1), 38–43.

Littig, B., & Grießler, E. (2005). Social sustainability: a catchword between political pragmatism and social theory. *International Journal of Sustainable Development*, *8* (1), 65-79.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lozano, R. (2006). Incorporation and institutionalization of SD into universities: breaking through barriers to change. *Journal of Cleaner Production*, *14*(9), 787-796. doi: 10.1016/j.jclepro.2005.12.010.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate behavioral research*, *33*(2), 181-220.

Marshall, J. D., & Toffel, M. W. (2005). Framing the elusive concept of sustainability: a sustainability hierarchy. *Environmental science & technology*, *39*(3), 673-682.

Matsuura, K. (2007). *Address on the occasion of the Round Table on ''Education and Economic Development''* UNESCO. Retrieved from: http://unesdoc.unesco.org/images/0015/001541/154152e.pdf

McPeck, J. E. (1981). *Critical thinking and education.* New York: St. Martins Press.

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*(4), 10–12.

Meadows, D., D. Meadows, and J. Randers. (1992). *Beyond the limits: global collapse or a sustainable future.* London: Earthscan Publications.

Miller, R., & Leskes, A. (2005). *Levels of assessment: From the student to the institution*. Retrieved from https://www.aacu.org/sites/default/files/files/publications/LevelsOfAssessment.pdf

Moore, J. (2005). Is higher education ready for transformative learning? A question explored in the study of sustainability. *Journal of Transformative Education*, *3*(1), 76-91.

Munda, G. (1997). Environmental economics, ecological economics, and the concept of sustainable development. *Environmental values*, *6*(2), 213-233.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*, *75*, 1-45.

National Commission on Excellence in Education. (1983). *A nation at risk : the imperative for educational reform : a report to the Nation and the Secretary of Education, United States Department of Education*. Washington, D.C.: National Commission on Excellence in Education: Supt. of Docs., U.S. G.P.O. distributor.

Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Omann, I., & Spangenberg, J. H. (2002). *Assessing social sustainability*. Paper presented at the biennial conference of the International Society for Ecological Economics, Tunisia.

Organization for Economic Cooperation and Development [OECD]. (2013) *OECD Factbook 2013: Economic, Environmental and Social Statistics.* Retrieved from http://www.oecd-ilibrary.org/economics/oecd-factbook-2013_factbook-2013-en

Ozturk, C., Dicle, A., & Muslu, G. K. (2008). A comparison of problem-based and traditional education on nursing students' critical thinking dispositions. *Nurse Education Today, 28*(5), 627-632. doi: 10.1016/j.nedt.2007.10.001

Paul, R. W. (1981).Teaching critical thinking in the ‘strong’ sense: A focus on self deception, world views, and a dialectical mode of analysis. *Informal Logic, 4*(2), 2-7.

Pezzey, J., & Toman, M. A. (2002). *The economics of sustainability: a review of journal articles*. Washington, DC: Resources for the Future.

Ponting, C., (2007). *A New Green History of the World: The Environment and the Collapse of Great Civilizations*. New York: Penguin Books.

Quist, J., Rammelt, C., Overschie, M., & de Werk, G. (2006). Backcasting for sustainability in engineering education: the case of Delft University of Technology. *Journal of Cleaner Production*, *14*(9), 868-876. doi:10.1016/j.jclepro.2005.11.032.

QSR International. (2014). NVivo [Computer software]. Retrieved from http://www.qsrinternational.com/about-qsr.aspx.

Reckase, M. (2009). *Multidimensional item response theory*. New York

Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers, & R. D. Hays (Eds.), *Assessing Quality of Life in Clinical Trials: Methods of Practice* (2nd ed.) (pp. 55-73). Oxford, NY: Oxford University Press.

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. *Current directions in psychological science*, *14*(2), 95-101.

Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, *3*(2), 85-93. doi: http://dx.doi.org/10.1016/j.tsc.2008.03.005

Robinson, J., 2004. Squaring the circle? Some thoughts on the idea of sustainable development. *Ecological Economics 48*, 369–384.

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3-13. doi: 10.1111/j.1745-3992.2005.00006.x

Roznowski, M. (1989). Examination of the measurement properties of the Job Descriptive Index with experimental items. *Journal of Applied Psychology*,*74*(5), 805-814. doi: 10.1037/0021-9010.74.5.805

Saravanamuthu, K., Brooke, Carole., & Gaffikin, M. (2013). ERP and precautionary ethics: harnessing critical thinking to engender sustainability. *Journal of Information, Communication and Ethics in Society, 11*(2), 92-111. doi: 10.1108/jices-12-2012-0020

Sireci, S. G., & et al. (1991). On the Reliability of Testlet-Based Tests. *Journal of Educational Measurement*, *28*(3), 237-247.

Serageldin I. (1993). Making development sustainable. *Finance & Development.* 30(4):6- 10

Sosu, E. M. (2013). The development and psychometric validation of a Critical Thinking Disposition Scale. *Thinking Skills and Creativity*, *9*, 107-119. doi: 10.1016/j.tsc.2012.09.002

Stavins, R. N., Wagner, A. F., & Wagner, G. (2003). Interpreting sustainability in economic terms: dynamic efficiency plus intergenerational equity. *Economics Letters*, *79*(3), 339-343. doi: 10.1016/S0165-1765(03)00036-3.

Steinberg, L., & Thissen, D. (1996). Uses of Item Response Theory and the Testlet Concept in the Measurement of Psychopathology. *Psychological Methods, 1*(1), 81-97. doi: 10.1037/1082-989x.1.1.81

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Thousand Oaks: Sage Publications, Inc.

Sternberg, R. J. (1986). *Critical Thinking: Its Nature, Measurement, and Improvement.* Washington D.C.: U.S. Department of Education

Swaminathan, H., & Gifford, J. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, *50*(3), 349-364. doi: 10.1007/bf02294110

Tabachnick, B. G. (2007). *Using multivariate statistics* (5th ed). Boston: Boston : Pearson/Allyn & Bacon.

Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students'critical thinking skills. *Research in Higher Education, 36*, 23-39.

Thorpe, G. L., & Favia, A. (2012). Data Analysis Using Item Response Theory Methodology: An Introduction to Selected Programs and Applications. *Psychology Faculty Scholarship.* Paper 20.

Thomas, I. (2009). Critical thinking, transformative learning, sustainable education, and problem-based learning in universities. *Journal of Transformative Education*, *7*(3), 245-264. doi: 10.1177/1541344610385753

Tilbury, D. (1995). Environmental education for sustainability: Defining the new focus of environmental education in the 1990s. *Environmental Education Research*, *1*(2), 195-212. doi:10.1080/1350462950010206

Tilbury, D. (2004). Rising to the challenge: education for sustainability in Australia. *Australian Journal of Environmental Education*, *20*(2), 103-114.

Tishman, S. (1994). What makes a good thinker: A look at thinking dispositions. *Harvard Graduate School of Education Alumni Bulletin*, *39*(1), 7-9.

Toman, M.A. (2006). The difficulty in defining sustainability. In, *The RFF reader in environmental and resource policy*. (pp. 1-6). Washington, DC: Resources for the Future.

Toutkoushian, R. K. (2005). What can institutional research do to help colleges meet the workforce needs of states and nations? *Research in Higher Education, 46*, 955-984. doi: 10.1007/s11162-005-6935-5

Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. doi: 10.1007/bf02291170

Turner, J. C. (1995). The influence of classroom contexts on young children’s motivation for literacy. *Reading Research Quarterly, 30*(3), 410–441.

Tullos, D. (2009). Assessing the influence of environmental impact assessments on science and policy: An analysis of the Three Gorges Project. *Journal of Environmental Management, 90*(3), S208-S223. doi:10.1016/j.jenvman.2008.07.031

University Leaders for a Sustainable Future [ULSF]. (1990). *The Talloires Declaration.* Retrieved from University Leaders for a Sustainable Future website: http://www.ulsf.org/programs_talloires.html

United Nations (Ed.). (1993). *Earth Summit: Agenda 21, the United Nations programme of action from Rio*. New York: United Nations.

Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching, 53*(1), 41–48.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Fam Med, 37*(5), 360-363.

Waas, T., Hugé, J., Verbruggen, A., & Wright, T. (2011). Sustainable development: A bird's eye view. *Sustainability, 3*(10), 1637-1661.

Wals, A. E., & Jickling, B. (2002). "Sustainability" in higher education: from doublethink and newspeak to critical thinking and meaningful learning. *International Journal of Sustainability in Higher Education*, *3*(3), 221-232.

Wagner, A. and Dobrowolski, J.W. (Eds). (2000). *International Conference on Holistic Concepts of Training for the Promotion of Sustainable Development Focused on European Integration for Better Quality of Environment and Human Life*, Book of Summaries. Krakow: University of Mining and Metallurgy.

Watson, G. B., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal*. San Antonio: The Psychological Corporation.

Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach? *American Educator, 109*(4), 21-32. doi: 10.3200/AEPR.109.4.21-32

World Commission on Environment and Development [WCED].(1987). *Our Common Future*, Oxford: Oxford University Press.

Worldwatch Institute. (2013). *Use and capacity of global hydropower increases.* Retrieved from http://www.worldwatch. org/node/9527.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. The Counseling Psychologist, 34(6), 806-838. doi:10.1177/0011000006288127

APPENDICES

Appendix A Sustainability Scales

Author	Title	Purpose	Population	α
Alcock, I (2012)	Measuring Commitment to Environmental Sustainability	To investigate the relationship between commitment to environmental sustainability and education	UK nationals 16+ years	.67
Maloney, M. P., & Ward, M. P. (1973)	Ecology: Let's Hear from the People	To understand what people know about ecology	College and Non-college students	SS1 =.91 SS2 = .93 SS3 = .92 SS4 = .89
Kagawa, F. (2007)	Dissonance in students' perceptions of sustainable development and sustainability	Students' understandings perceptions of sustainability related concepts	Undergraduate students	None provided
Azapagic, A., Perdan, S., & Shallcross, D. (2005)	How much do engineering students know about sustainable development?	Engineering students' knowledge of sustainable development	Undergraduate Engineering Students	None provided
Holl, K. D., Daily, G. C., Daily, S. C., Ehrlich, P. R., Bassin, S. (1999)	Knowledge of and attitudes toward population growth and the environment	To understand students' knowledge of, and attitudes towards, population growth and the environment	Undergraduate students	None provided

Note. SS is Sub-scale

Appendix B Critical Thinking Scales

Author	Title	Purpose	Population	α
Insight Assessment (2013)	California Critical Thinking Tests (CCTSS)	To measure critical thinking skills for educational or workplace settings	Undergraduate Graduate	None provided
ACT (2014)	Collegiate Assessment of Academic Proficiency (CAAP)	To evaluate student learning general education program outcomes.	Freshman and Sophomore	None provided
Ennis and Weir (1985)	Ennis-Wier Critical Thinking Essay Test	To measure critical thinking	High school and College	None provided
Watson and Glaser	Watson-Glaser Critical Thinking Appraisal	Measures cognitive abilities in professionals	Graduate, professionals and managers	None provided
Council for Aid to Education (CAE) (2014)	Collegiate Learning Assessment (CLA)	Measures an institution's contribution to student development and learning	Undergraduate	None provided
Tennessee Tech University (2014)	Critical Thinking Assessment Test(CAT)	To assess and promote the improvement of critical thinking and real-world problem solving skills	Undergraduate	None provided

Note. α indicates scale reliability

Appendix C Interview Protocol

Part 1: Background information about dissertation, instructor, and course

Thank you very much for allowing me to interview you, your participation in this study very much appreciated. Today is __/__/__ and my name is Vivian Alexander. I am a PhD candidate in educational psychology, and I am currently working on my dissertation research. The purpose of this research is to develop and validate a scale to evaluate students' ability to think critically about sustainability. I am utilizing a sequential mixed methods design that consists of two phases: Phase 1 involves one-to-one interviews with experts on sustainability; and Phase 2 a quantitative phase in which I use the information from these interviews to help guide my research and the development of the scale.

A. Instructor background

First, I'd like to ask you a few questions about yourself.

1. Please state your name and your position at this university
2. How long have you been teaching/doing research at the college or university level?
3. How many years have you been teaching/conducting research at this university?

B. Sustainability: Now I am going to ask you some questions about sustainability.

1) **What does the term "sustainability" mean to you?**
   Follow-up question(s) if necessary:
   a. What about economic sustainability?
   b. What about social sustainability?
   c. What about ecological sustainability?
2) **In your opinion, is there a difference between sustainability and sustainable development?**
   a. Why or why not?
   b. What differentiates the two?
3) **In your experience, what do students know about sustainability?**
   a. What are students' attitudes about sustainability?
      i. Are they generally interested or uninterested?
      ii. Do you notice any difference in students' attitudes about sustainability based on their class standing: freshman, sophomore, junior, and senior?
   b. Why is important for students to know, be aware of, or show concern for sustainability.
4) **How would you or how have you measure(d) students' knowledge of sustainability?**
   a. In your opinion, is the ability to think critically a part of or important for learning about sustainability?

Appendix D Permissions

Dear Vivian--
The permission you requested is granted below, provided that the cases are posted in a password environment that users must log in to. We do not allow reposting of any of our material on public, open access sites.
Ky
~~~~~~~~~~~

As the copyright holder, the National Center for Case Study Teaching in Science is pleased to grant permission to Vivian Alexander to adapt the following material for use within a password-protected online system restricted to participants surveyed for her Dissertation work carried out at Purdue University. This permission extends only to the material identified below and does NOT include teaching notes or any answer key material. Permission is granted as of January 8, 2015 and extends for two years from that date, after which time the material must be removed or a request for extension must be made.

1) Ecotourism-Who Benefits
2) Watch Your Step: Understanding the Impact of Your Personal Consumption on the Environment
3) A Struggle for Power in China: Three gorges dam
4) The Wealth of Water: The Value of an Essential Resource
5)To Be or Not To Be a Golf Course in Wimberley?

Please note that this permission only covers textual content and images for which we retain copyright and does not cover the use of additional images that may appear. We are unable to grant permission for their usage because they are either in the public domain, or were licensed or used with permission from other copyright holders (this usually only affects images used for design purposes--their credit information is generally found at the bottom of the pages on which they appear or at the end of the document). As a publisher we are bound by certain conditions that may not apply to your instructional use. You will need to make your own judgment regarding whether further permission is required for your situation or whether 'fair use' guidelines apply.

Our conditions are the following:
(i) That the title block information be retained so that the original author of a case is clearly acknowledged (formatting and placement may be changed to your liking); where applicable, you may wish to note that the case has been adapted.
(ii) That you append the following statement somewhere in association with the case: "Copyright held by the National Center for Case Study Teaching in Science, University at Buffalo, State University of New York, all rights reserved. Used with permission."
~~~~~~~~~~~

We are always interested in getting feedback from people who use our cases. Please use the comments tab on the case record pages to tell us about your experience with the material.

Please let me know if you have any questions.

Sincerely,
Ky Herreid
Permissions Manager at NCCSTS
National Center for Case Study Teaching in Science
University at Buffalo Biological Sciences
109 Cooke Hall
Buffalo, NY 14260
permissions@sciencecases.org

~~~ORIGINAL REQUEST~~~
Profit Status: Not-for-Profit
License: Other
Web Address: N/A
Cases: 1) Ecotourism-Who Benefits
2) Watch Your Step: Understanding the Impact of Your Personal Consumption on the Environment
3) A Struggle for Power in China: Three gorges dam
4) The Wealth of Water: The Value of an Essential Resource
5)To Be or Not To Be a Golf Course in Wimberley?
Course: Dissertation
Semesters: Spring 2015
Enrollment: Over 200
Comments: Dear Sir/Madam,

My name is Vivian Alexander and I am a Ph.D. candidate in the Educational Studies Department at Purdue University. My dissertation research involves the creation of a scale to measure undergraduate engineering students' ability to think critically about sustainability issues. I found a number of very interesting and useful cases studies that I would like to use/adapt for my dissertation research. However, my potential sample size requires that I apply for formal permission.

I will be administering the questions online via qualtrics survey software.

First Name: Vivian Last Name: Alexander
Department: Educational Studies Institution: Purdue University
City: West Lafayette Telephone: 7655320650
Email: valexand@purdue.edu Agreement: I agree not to reproduce teaching notes or answer keys.
~~~

Appendix E Scenario 1

The Ogallala Aquifer, a groundwater formation that underlies much of the Great Plains, got the nation's attention when Nebraska landowners and environmentalists protested the shipment of oil from Canada's tar sands to the Texas coast, saying that a leak in the proposed Keystone XL pipeline might damage the aquifer. They won when the State Department announced it would reconsider the pipeline's route.

The center-pivot irrigation systems that have turned the ground all the way from South Dakota to Texas into a giant mesh of green, gold, and brown circles spray Ogallala water twenty-four hours per day throughout much of the summer— and up to six trillion gallons of water per year. Unfortunately, Ogallala water is not self-renewing. Geologists call it fossil water because it took thousands of years to collect underground. However, farmers and big ag-industrial companies argue that the water is put to good use and without it, many people would go hungry. However, in just 70 years, irrigators have run out of water in many places. They will run out in most other areas well before the end of this century.

The problem worsens as federal policies that encourage the growth of one of the thirstiest crop—corn in the region, only serve to increase the rate of decline of the Ogallalla Aquifer. An ethanol mandate still in place ensures that, by 2015, over one-third of the nation's corn will become fuel. Most of the rest becomes corn syrup or is fed to cattle and turned into fatty beef. Neither of these foods is good for us, as the nation's heart disease and diabetes epidemics testify. Moreover, United States Geological Survey studies show that agriculture chemicals are showing up in the water. Therefore, the federal Farm Program is helping to destroy the Ogallala aquifer and to sicken the nation by giving support payments to corn farmers regardless of how many chemicals or how much water they use.

Questions:

1. Describe the sustainability issues and challenges
2. Identify additional information needed to thoroughly consider the sustainability issues
3. Recommend a strategy for resolving the sustainability issue. How or why would this strategy resolve the sustainability issue?
4. Describe the implications and consequences of your recommendation.
5. Who are the key stakeholders? In what ways are their interest aligned or in competition?
6. Describe how your own knowledge, perspectives, and opinions influence your interpretation of this case study.
7. What does the term "Sustainability" mean to you?

Appendix F Scenario 2

The human diet has various components: grains, meat, dairy, fruit and vegetables. Fresh foods are diverse elements of our diet that complement grains but have different characteristics than require preservation. Unlike grains, which have a long shelf life, perishable produce such as dairy spoils easily. A third of the world's food goes to waste. In the developed markets, about one-third of our food is lost because we buy too much and we throw it away. In the developing economy, it is a different problem. The food never makes it out of harvest. It rots on the field because there is not a good transportation infrastructure, or if it is transported it is transported in poor conditions to a wet market, an outdoor market, where the food rots waiting for consumers to buy it. To support a growing population there is need to extend the supply of foods which are not making it to our tables due to food spoilage and waste.

We produce enough food on the planet today to feed everybody, and if we do it more efficiently—we feed the people tomorrow. We just need to make sure that we get the food that we make to the place that it needs to go. The average piece of produce in the U.S. travels 1,500 miles from its source contributing to the huge carbon footprint of food production. Much of the wastage is due to the lack of proper refrigeration, especially in the developing world where as much as a third of all of the food produced by farmers goes to waste, most before it even gets into a kitchen. Yet millions of people do not get enough to eat, and the carbon footprint of all that wasted food is enormous. So, it's an inefficient system in both developed and developing economies. Food waste by itself represents 3.3 billion metric tons of carbon dioxide. If you measured food loss as a country on its own, it would be the third largest emitter of greenhouse gases behind China and the United States.

Questions:

1. Describe the sustainability issues and challenges
2. Identify additional information needed to thoroughly consider the sustainability issues
3. Recommend a strategy for resolving the sustainability issue. How or why would this strategy resolve the sustainability issue?
4. Describe the implications and consequences of your recommendation.
5. Who are the key stakeholders? In what ways are their interest aligned or in competition?
6. Describe how your own knowledge, perspectives, and opinions influence your interpretation of this case study.
7. What does the term "Sustainability" mean to you?

Appendix G Scenario 3

Recent legislation in the United States and Europe has produced policies that limit or ban the sale of incandescent light bulbs. These policies were established to increase the efficiency of energy use in building lighting. Compact fluorescent light (CFL) bulbs were the first alternative technology widely available in the market place. More recently light emitting diode (LED) bulbs have become readily available. CFLs and LEDs use approximately 79 and 88 percent less energy, respectively, than conventional incandescent lighting. Widespread use of more efficient lighting leads to lower energy use and a reduction in greenhouse gas (GHG) emissions. A single CFL or LED bulb is significantly more expensive than an equivalent illumination incandescent bulb.

However, CFL bulbs last about 10 times longer, and LED bulbs 50 times longer, than incandescent bulbs. Thus, over the lifetime of a CFL or LED bulb the replacement cost of incandescent bulbs and the higher energy use of incandescent illumination make CFL and LED bulbs a far superior economic value. Public perception of the light quality is another issue slowing the adoption of more efficient lighting. As technology advances, designers are learning to tune the light color to satisfy different market demands.

CFL and LED bulbs contain materials that are toxic and/or scarce. In particular, CFL bulbs contain mercury that is released if the bulb is broken. The environmental protection agency (EPA) determined that mercury in CFL bulbs did not pose a serious hazard to consumers. The largest source of mercury emissions to the environment is from burning coal for electricity generation. Despite CFL bulbs containing mercury, using them for lighting results in far lower emission of mercury to the environment relative to incandescent bulbs due to reduce demand for electricity. There are known tradeoffs for using new energy efficient lighting. For an equivalent service period of illumination, CFL and LED bulbs yield substantial (>80%) reductions in primary energy demand and GHG emissions. However, CFL and LED bulbs have higher human and eco-toxicity potential and resource depletion potential than incandescent bulbs.

Questions

1. Describe the sustainability issues and challenges
2. Identify additional information needed to thoroughly consider the sustainability issues
3. Recommend a strategy for resolving the sustainability issue. How or why would this strategy resolve the sustainability issue?
4. Describe the implications and consequences of your recommendation.
5. Who are the key stakeholders? In what ways are their interest aligned or in competition?
6. Describe how your own knowledge, perspectives, and opinions influence your interpretation of this case study.
7. What does the term “Sustainability” mean to you?

Appendix H Critical Thinking about Sustainability Rubric

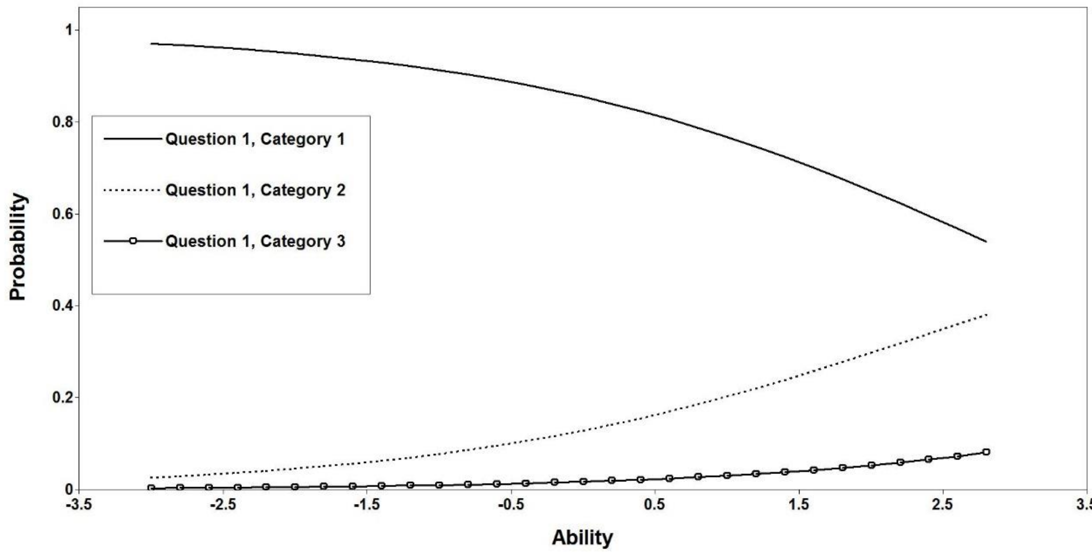
	Proficient	Emerging	Developing
Q1: Describe the sustainability issues and challenges and explain the pros and cons of these issues or challenges.	The student fully describes the dimensions of ecological, economic, and social issues/challenges.	The student fully describes two of the three sustainability challenges	The student describes only one sustainability or challenge
Q2: Identify additional information needed to thoroughly consider the sustainability issues	The student identifies at least three additional pieces of information necessary to thoroughly consider the sustainability issues	The student identifies at 1-2 pieces of information necessary to thoroughly consider the sustainability issues	The student identifies 0-1 piece of information necessary to thoroughly consider the sustainability issues
Q3: Recommend a strategy for resolving the sustainability issue.	The recommended strategy is plausible and student clearly describes how or why the recommended strategy would resolve the sustainability issue	The recommended strategy is plausible but the student's description of how or why the recommended strategy would resolve the sustainability issue is lacking.	The recommended strategy is not plausible and the student does not clearly describe how or why the recommended strategy would resolve the sustainability issue
Q4: Describe the implications or consequences of your recommendation.	The student clearly describes the implications or consequences of their recommendations	The students' description of the implications of their recommended strategy is unclear or difficult to understand.	The students' descriptions of the implications of their recommended strategy is weak or non-existent

Appendix H continued

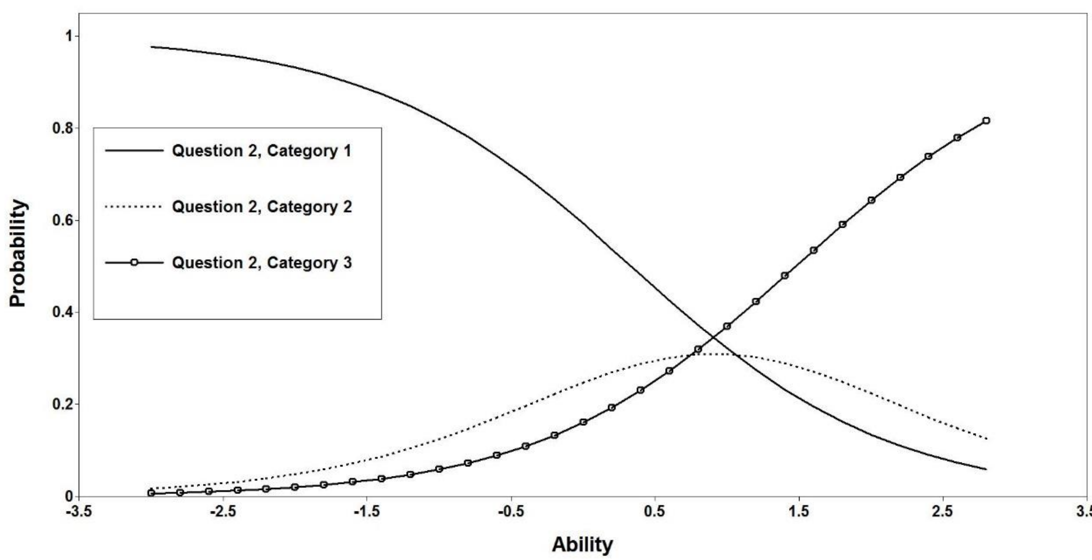
Q5: Identify all the stakeholders involved and how their interests are aligned or in competition	The student identifies all stakeholders and clearly describes how their interests are aligned or in competition	The students identifies all stakeholders, but lacks a clear description of how their interests are aligned or in competition	The student does not identify all key stakeholders. ***<u>Or,</u>*** The student identifies some key stakeholders, but fully describes how their interests are aligned or in competition
Q6: Describe how your own knowledge, perspectives, and opinions influence interpretation of this case study.	The student describes in detail how their own knowledge, perspectives, and opinions influence their interpretation of this case study.	The students provides some detail about how their own knowledge, perspectives and opinions influence their interpretation of this case study	The students ***<u>does not</u>*** describe in detail how their own knowledge, perspectives and opinions influence their interpretation of this case study

Note. Q = Question. Q7 note included because it only asks students what sustainability means to them. The question is not a measure of students CT ability.

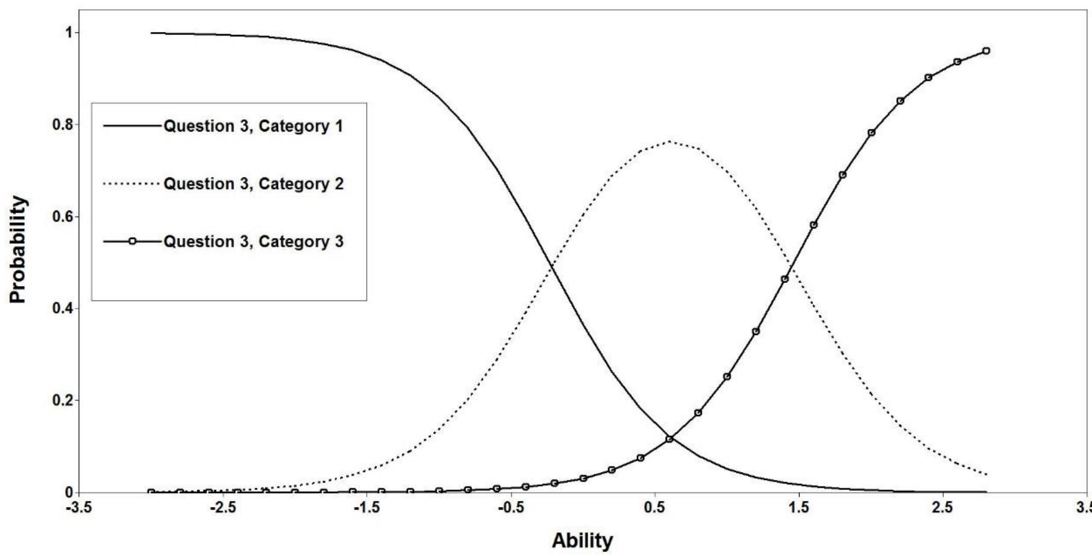
Appendix I Category Characteristic Curves for KFCC Scenario



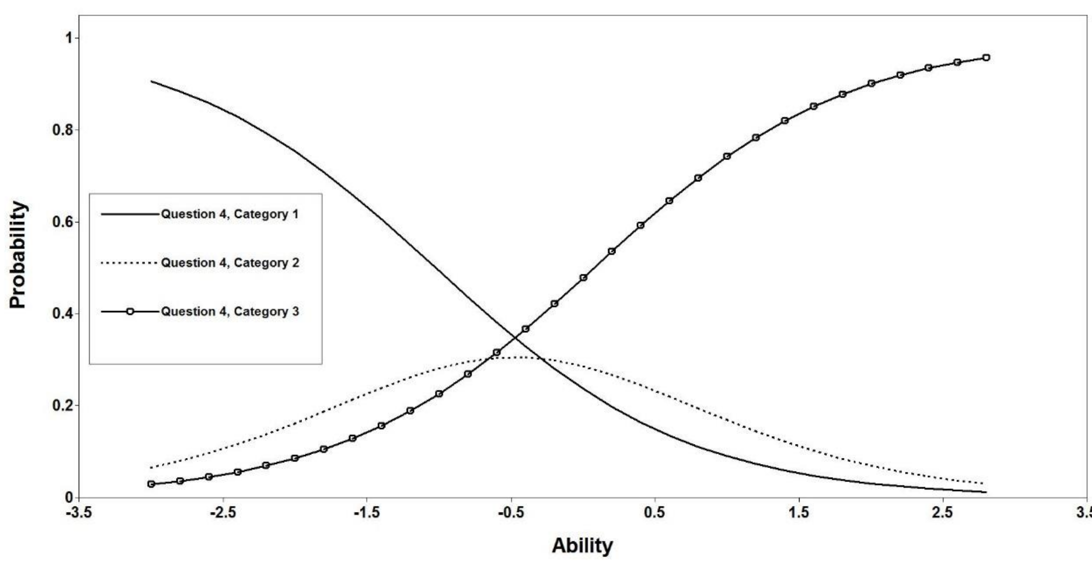
Category Characteristic Curve KFCC Question 1



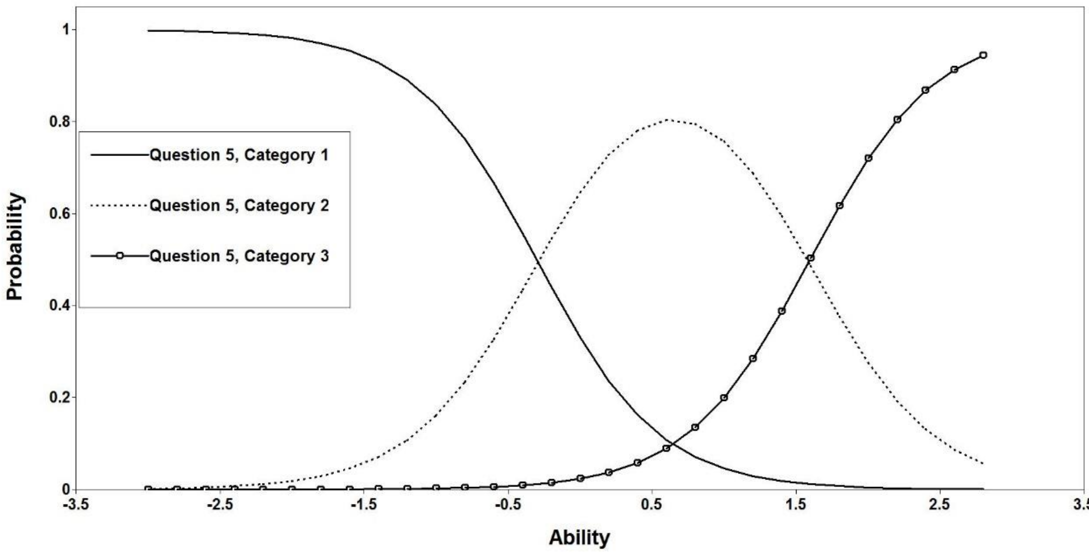
Category Characteristic Curve KFCC Question 2



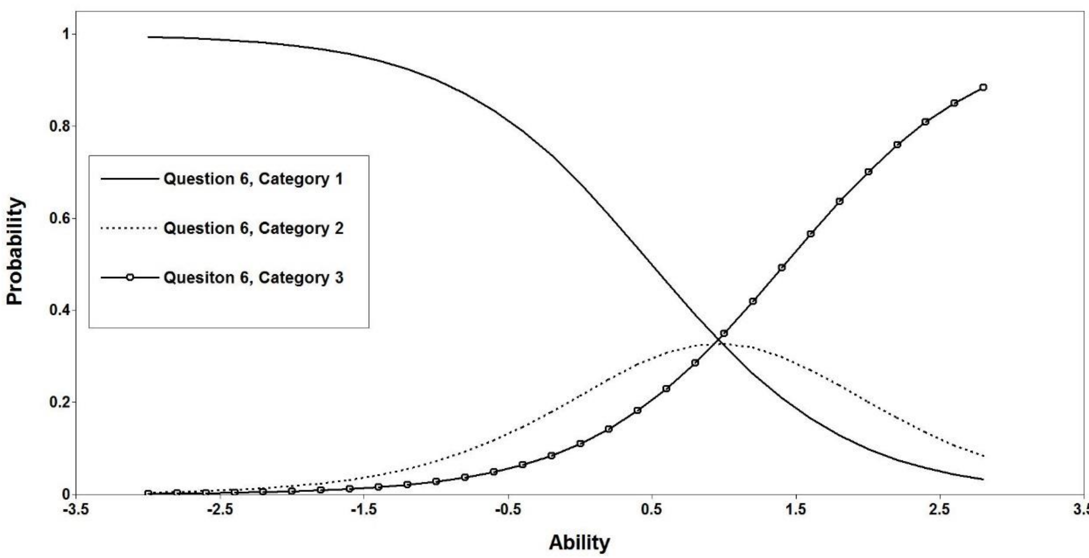
Category Characteristic Curve KFCC Question 3



Category Characteristic Curve KFCC Question 4

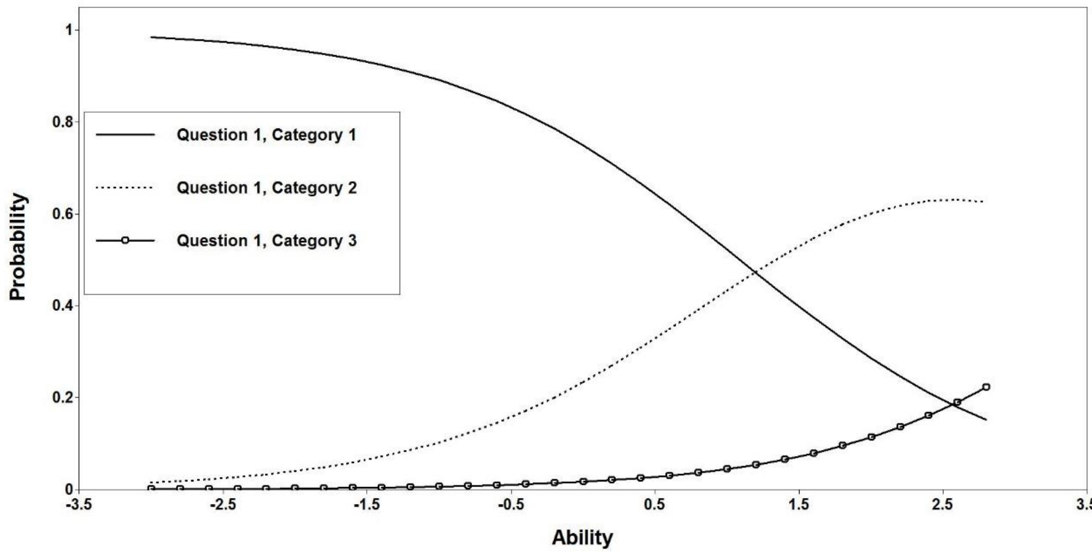


Category Characteristic Curve KFCC Question 5

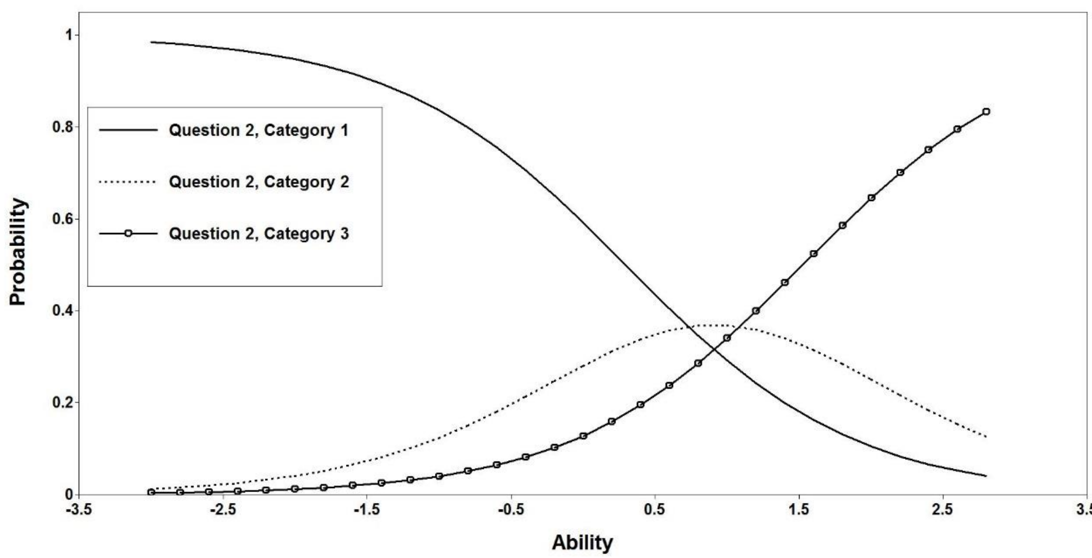


Category Characteristic Curve KFCC Question 6

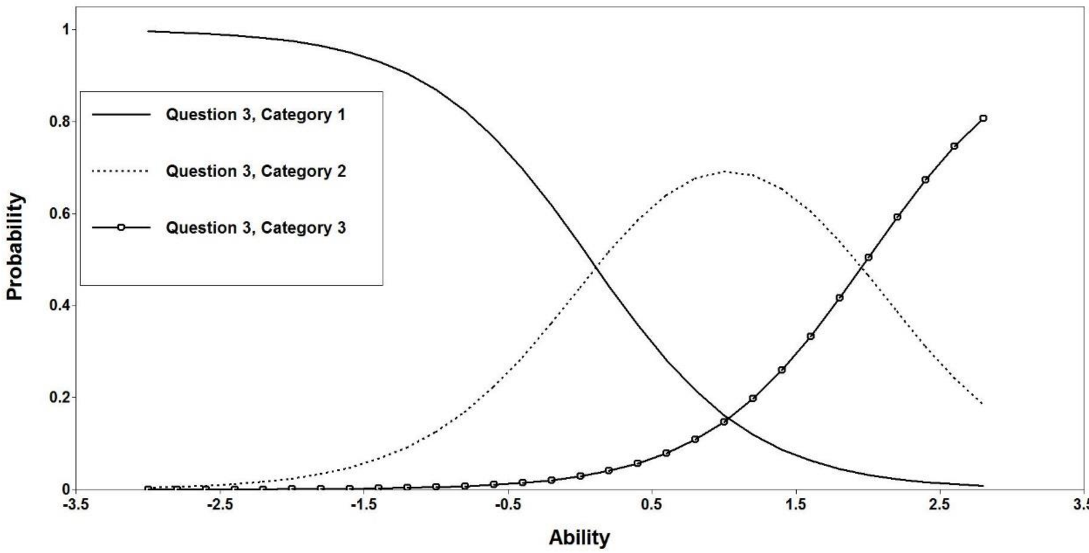
Appendix J Category Characteristic Curves for BBD Scenario



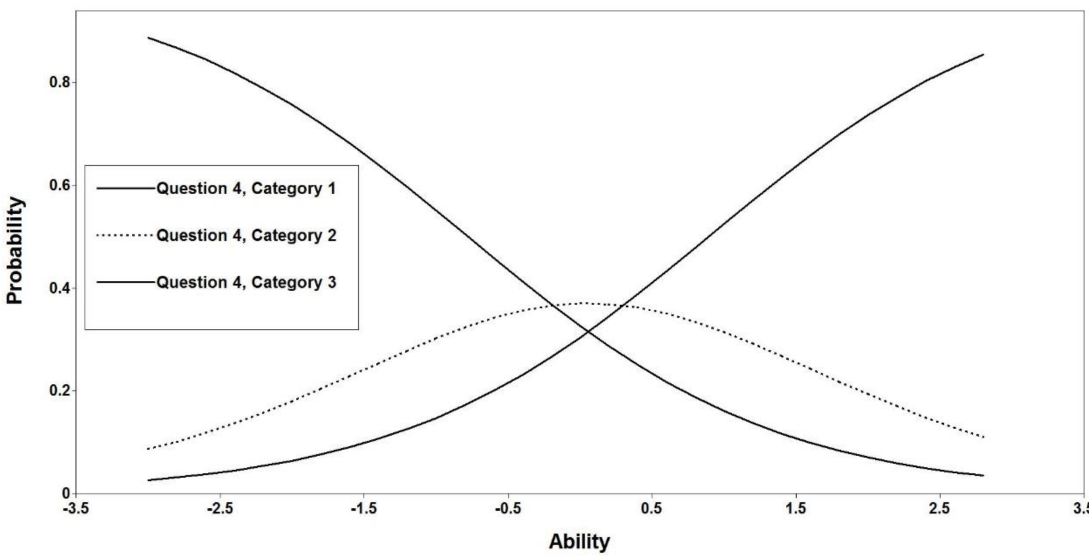
Category Characteristic Curve BBD Question 1



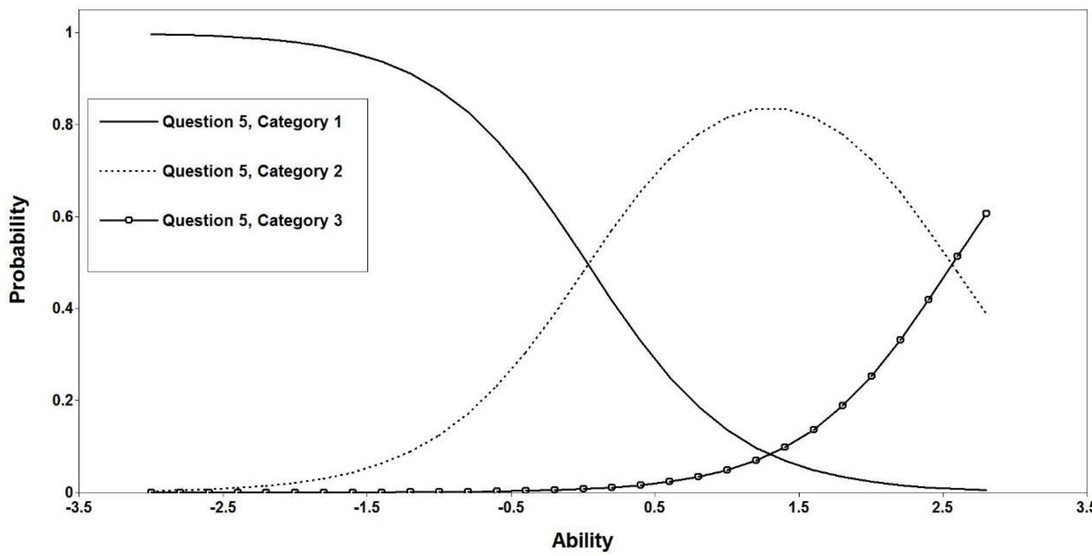
Category Characteristic Curve BBD Question 2



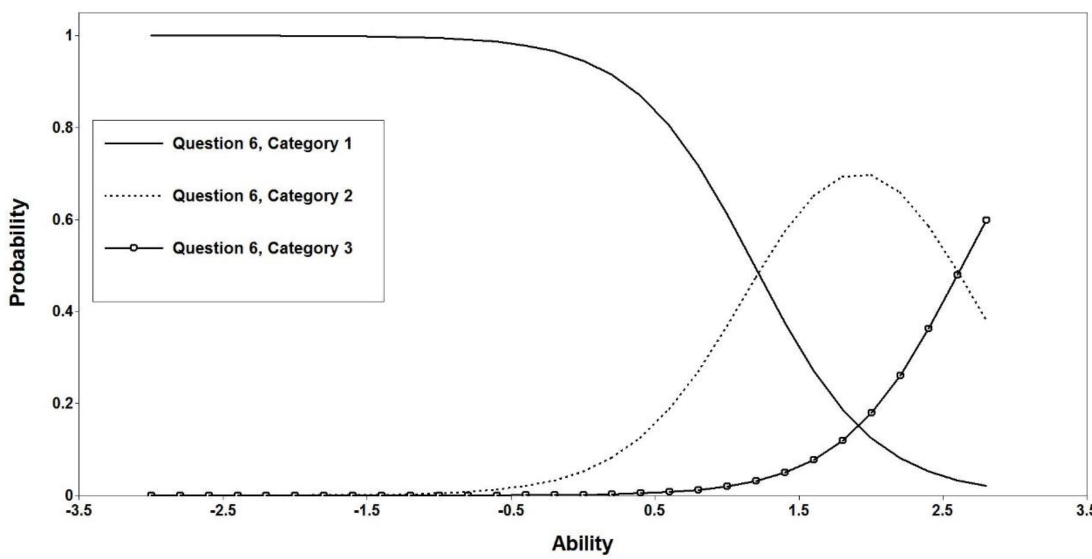
Category Characteristic Curve BBD Question 3



Category Characteristic Curve BBD Question 4

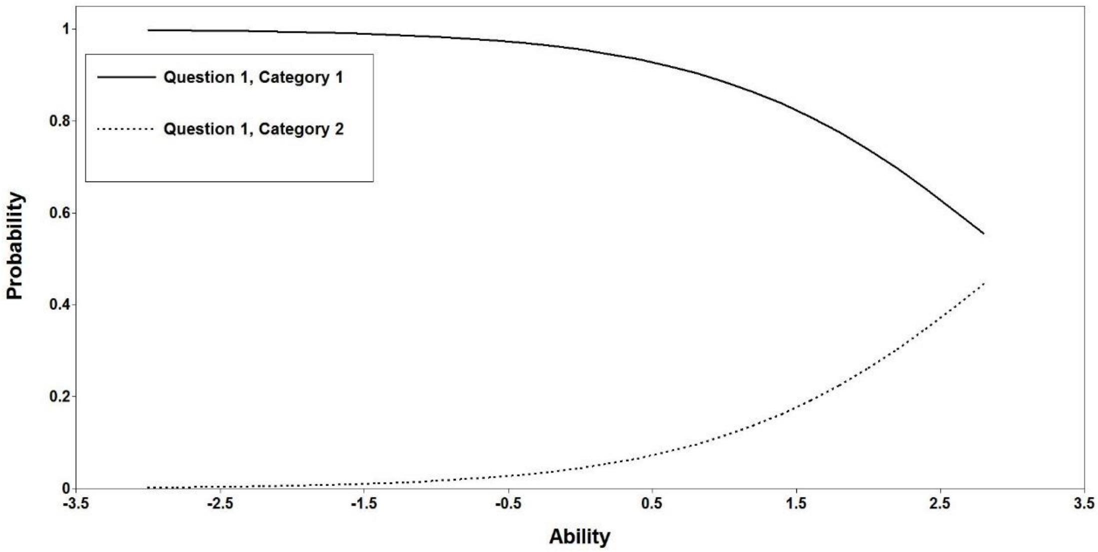


Category Characteristic Curve BBD Question 5

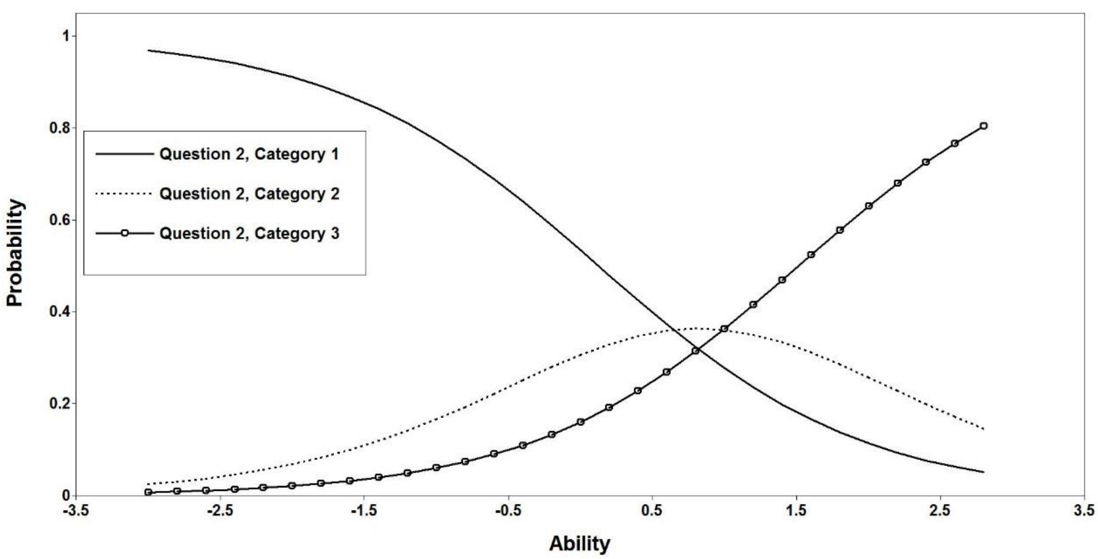


Category Characteristic Curve BBD Question 6

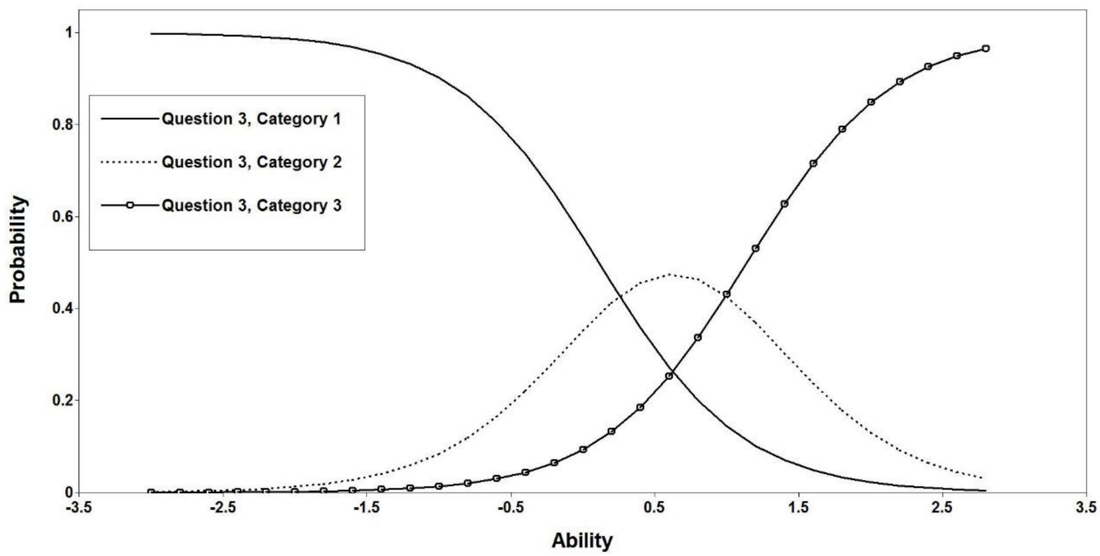
Appendix K Category Characteristic Curve for OTO



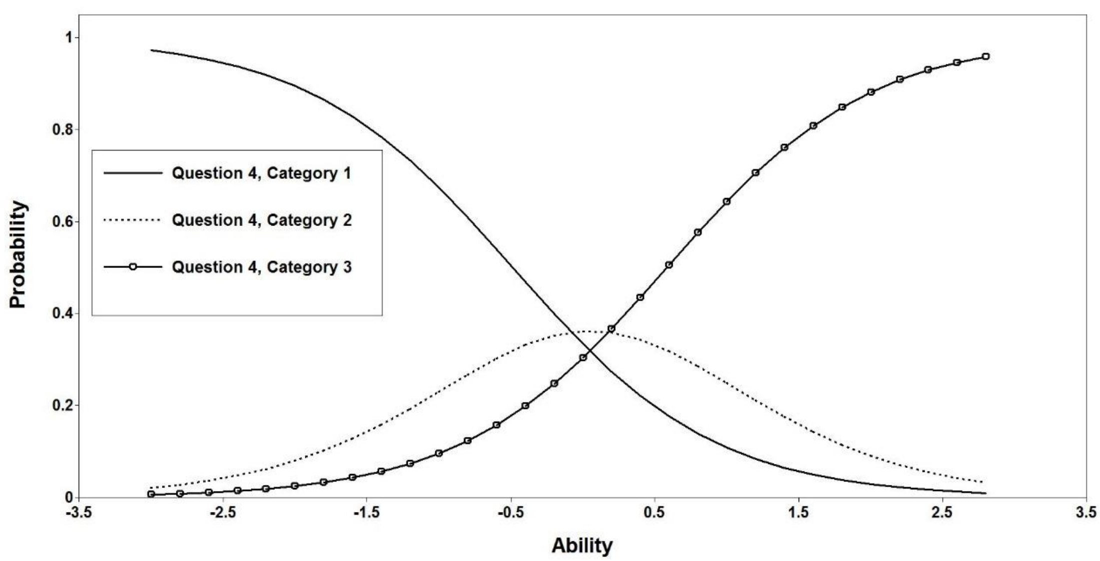
Category Characteristic Curve OTO Question 1



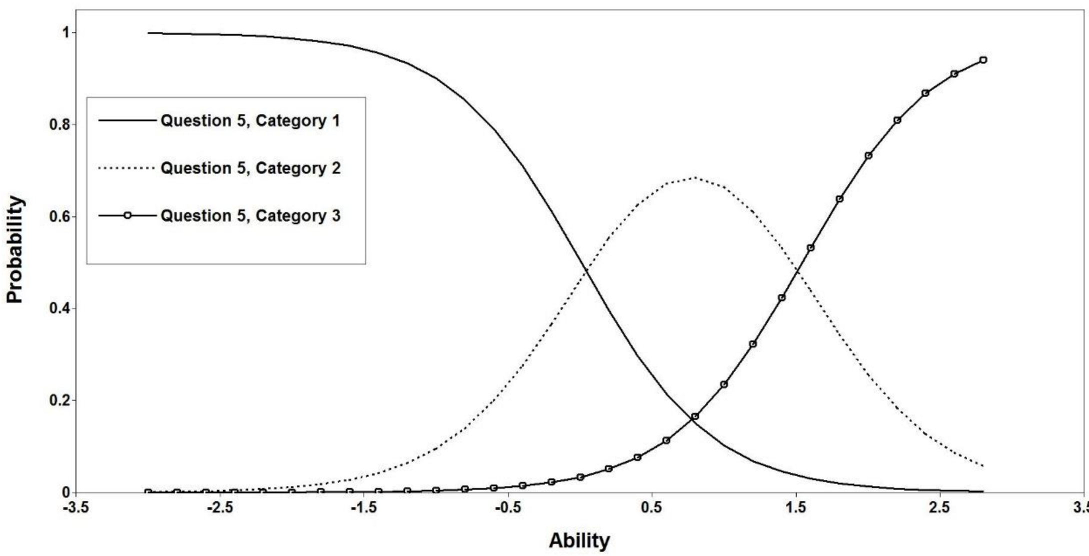
Category Characteristic Curve OTO Question 2



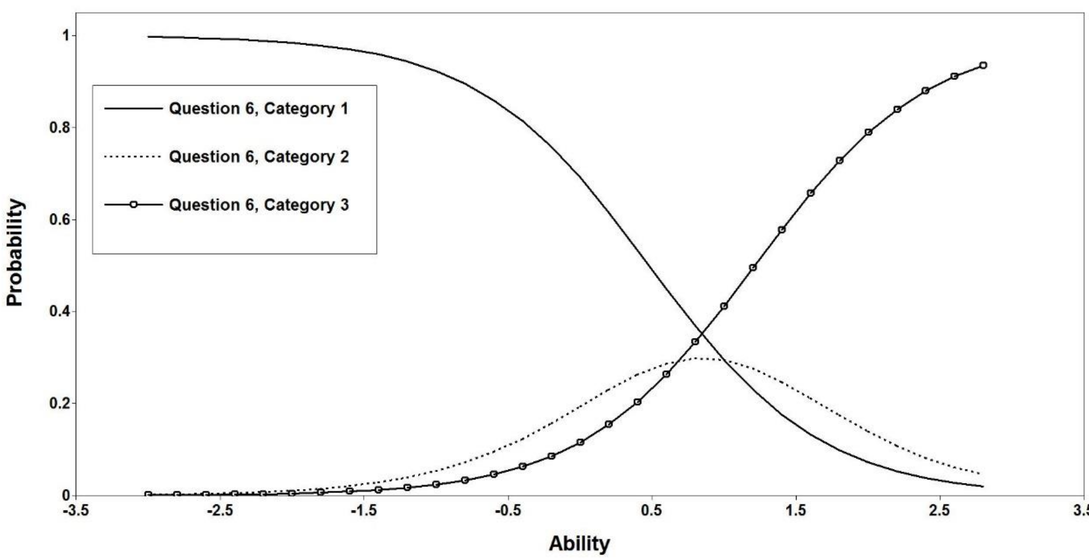
Category Characteristic Curve OTO Question 3



Category Characteristic Curve OTO Question 4



Category Characteristic Curve OTO Question 5



Category Characteristic Curve OTO Question 6

VITA

VITA

VIVIAN GREGORY ALEXANDER

Ph.D. Candidate
Educational Psychology specializing in Applied Measurement and Research Methodology
Department of Educational Studies Purdue University
Advisor: Yukiko Maeda, Ph.D.

EDUCATION

Ph.D. (Expected Graduation: December, 2015) Development and Validation of the Thinking Critically about Sustainability Scale
B.Sc. Middle Level Education, Claflin University,Orangeburg, SC - May 2011
EMPHASIS: Middle School Mathematics and Science

RESEARCH INTERESTS

Methodological interests:
Hierarchical Linear Modeling
Factor Analysis
Item Response Theory

Substantive interests:
STEM achievement in Latin American and Caribbean Countries (Gender and equity issues)

FELLOWSHIPS, AWARDS AND HONORS

- Purdue Research Foundation (PRF) Grant, 2014 ($17, 232)
- Noeth Scholarship for Assessment and Evaluation, 2014 ($2,000)
- Noeth Scholarship for Assessment and Evaluation, 2013 ($2,000)
- Ross Fellowship, Purdue University, 2011 -2015
- Most Outstanding Student in Leadership and Excellence, Claflin University, School of Education, 2011
- President’s List, 2009, 2010
- Dean’s List, 2008
- Alice Carson Tisdale Honor’s College Scholarship, 2007

GRANTS

South Carolina Independent Colleges and Universities (SCICU) Faculty/Student Research Grant Amount: $3200 Date: February 2010.

The grant funded my research in my native country of Trinidad and Tobago, where I conducted the data collection portion of my research project titled "Working Memory and Writing from the perspective of a Developing nation." I worked under the supervision of Dr. Robert Vanderburg, Associate Professor of Education at Claflin University.

PRESENTATIONS (BY YEAR)

2015

Alexander, V. & Maeda, Y. (2015, June). Gender Differential Item Functioning in Trinidad and Tobago Students' Performance on PISA 2009. Presented at The University of the West Indies Biennial Conference, Barbados.

2014

Maeda, M., **Alexander, V.,** Senk, S.L., Craig, J., & He, J., (2014, April). *Preparation for Teaching Algebra: Results From a National Survey*. Paper presented at American Educational Research Association Annual Conference, Philadelphia.

2013

Alexander, V., & Maeda, M. (2013, April). *Student Achievement in Latin America and the Caribbean: Evidence from PISA 2009.* Presented at The University of the West Indies Biennial Conference, Trinidad and Tobago.

2012

Alexander, V., & Maeda, M. (2012, April). *Student Achievement in Reading Mathematics and Science in Latin America and the Caribbean: Evidence from PISA 2009.* Comparative International Education Society (CIES) Annual Conference, Puerto Rico.

Wortinger, K. Mintos, A**.,** **Alexander, V.,** & Maeda, Y. (2012, November). *Survey design for secondary mathematics teacher education programs: Challenge and Opportunities*. Paper presented at the 34th Annual Conference of the North American Chapter of the International Group for the Psychology of Mathematics Education (PME-NA 2012), Kalamazoo, MI.

2011

Alexander, V., & Vanderburg, R. (2011). *Working Memory and Writing from the Perspective of a Developing Nation.* South Carolina Independent Colleges and Universities Research Symposium, South Carolina.

2008

Alexander, V., & Vanderburg, R. (2008). *Working Memory and Transcription.* Asian Pacific Education Research Association (APERA) Bi-Annual Conference, Japan.

RESEARCH EXPERIENCE

Graduate Research Assistant:
College of Education, Purdue University, *Fall 2011 –Spring 2014*
Project:
NSF funded Project Preparing to Teach Algebra (PTA)
Principal Investigators (P.I.):
Dr. Sharon Senk (Michigan State University), Dr. Yukiko Maeda (Purdue University), Dr. Jill Newton (Purdue University)
Role:
Designing and implementing a national survey of 400 secondary teacher education programs

Research Consultant:
School of Engineering Education, Purdue University, *Fall 2014 – Spring 2014*
Project:
Pedagogical Evaluation Laboratory at Purdue
Principal Investigator (P.I.):
Dr. Monica Cox Sharon (Purdue University)
Role: Designing and validating the Global Real-Time Assessment Tool for Teaching Enhancement (G-RATE) for use in science, technology, engineering, and mathematics (STEM) classrooms.

Graduate Research Assistant:
College of Education, Purdue University, *Spring 2013 – Summer 2013*
Project:
Academic achievement of students from Latin American and Caribbean countries
Principal Investigator:
Dr. Yukiko Maeda (Purdue University)
Role:
Lead author on the manuscript due to submit for publication in May, and presented at the Biennial Conference of The University of the West Indies Schools of Education [April 23-25, 2013].

Graduate Research Assistant:
School of Engineering Education, Purdue University, *Summer 2013 - Fall 2013*
Project:
NSF funded project Information Literacy Skill Development and Assessment in Engineering (ILSDAE).

Principal Investigators (P.I.):
Dr. Senay Purzer (Purdue University), Dr. Michael Fosmire (Purdue University), Ms. Amy van Epps (Purdue University)
Role:
Designing and validating the Critical Engineering Literacy Test (CELT).

Graduate Research Assistant:
College of Education, Purdue University, ***Fall 2011 – Fall 2012***
Project:
Mathematics and Science achievement in Trinidad and Tobago: Evidence from PISA 2009
Principal Investigator (P.I.):
Dr. Yukiko Maeda (Purdue University)
Role:
Lead author on the manuscript and gave a brief presentation on the use of large-scale secondary data for conducting analyses at a seminar hosted by Career and Technical Education [Fall 2012].

INTERNSHIP EXPERIENCE

National Science Foundation, Directorate of Computer & Information Science & Engineering (CISE)

TEACHING EXPERIENCE

Student Teacher: Claflin University, Claflin University Upward Bound, 2008

Role: Preparing high school students to take the ACT and SAT examinations. Conducted pre- and post-tests to gauge their learning and a college search exercise with the students with the aim of identifying the necessary admissions requirements for the students' chosen colleges/universities

Teaching Assistant: Introduction to Education class- 2009 Claflin University

Role: Teacher's Assistant for an Introduction to Education class, designed to help students pass the Praxis 1 examination for teacher certification.

Teacher's Assistant: Student Support Services 2008 Claflin University

Role: Tutored students in Biology and Math

RELATED COURSES TAKEN

EDPS 632: Psychometric Theory and Applications, Fall 2014, Purdue University
EDPS 632: Structural Equation Modeling, Fall 2014, Purdue University
EDPS 632: Multilevel Modeling, Spring 2014, Purdue University
EDPS 638: Factor Analysis, Fall 2013, Purdue University
HDFS 627: Multilevel Modelling, Fall 2013, Purdue University
STATS 524: Multivariate Analysis, Fall 2012, Purdue University
STATS 512: Multiple Regression, Spring 2011, Purdue University

STATISTICAL SOFTWARE

SPSS, SAS, MPLUS, HLM6, BILOG-MG

PUBLICATIONS

PUBLICATIONS

Maeda, Y., **Alexander, V. G.,** Newton, J. & Senk S. L. (2014). Development of a National Survey for Secondary Mathematics Teacher Education Programs: Challenges and Lessons Learned. *Assessment Update*, *26*(3), 5-6, 12-13.

Newton. J., Maeda, Y., **Alexander, V.**, Senk, S., L. (2014). How well are secondary mathematics teacher education programs aligned with the recommendations made in *MET II*? *Notices of the American Mathematical Society*. *61*(3), 292-295.

Sambamurthy, N., Cox, M. F., **Alexander, V.,** Maeda, Y., & Perram, K. (2014, June) *Preliminary Analyses of Survey and Student Outcome Data using the Global Real-Time Tool for Teaching Enhancement (G-RATE).* American Society for Engineering Education (ASEE) Conference proceedings, Indianapolis, IN.

Wortinger, K., Mintos, A., **Alexander, V.,** & Maeda, Y. (2012, November). *Survey design for secondary mathematics teacher education programs: Challenge and Opportunities.* PME-NA 2012 Conference proceedings, Kalamazoo, MI.

Alexander, V. & Maeda, Y. (in press). The Complex Web: Mathematics and Science Achievement in Trinidad and Tobago, Evidence from PISA 2009. *Prospects, Quarterly Review of Comparative Education.*

Alexander, V. & Maeda, Y. (in press). Gender Differential Item Functioning in Trinidad and Tobago Students' Performance on PISA 2009. *Caribbean Education Research Journal.*